371 P

PATENT COOPERATION TREATY

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OF ELECTION

(PCT Rule 61.2)

To:

Assistant Commissioner for Patents
United States Patent and Trademark Office
Box PCT
Washington, D.C.20231
ETATS-UNIS D'AMERIQUE

in its capacity as elected Office

Date of mailing (day/month/year) 12 May 2000 (12.05.00)	
International application No. PCT/CA99/00852	**Applicant's or agent's file reference** CG/12326.19
International filing date (day/month/year) 15 September 1999 (15.09.99)	**Priority date** (day/month/year) 15 September 1998 (15.09.98)
Applicant ROUSSEAU, François	

1. The designated Office is hereby notified of its election made:

   [X] in the demand filed with the International Preliminary Examining Authority on:
   06 April 2000 (06.04.00)

   [ ] in a notice effecting later election filed with the International Bureau on:
   _____

2. The election [X] was

   [ ] was not

   made before the expiration of 19 months from the priority date or, where Rule 32 applies, within the time limit under Rule 32.2(b).

The International Bureau of WIPO
34, chemin des Colombettes
1211 Geneva 20, Switzerland

Facsimile No.: (41-22) 740.14.35

Authorized officer

Nestor Santesso

Telephone No.: (41-22) 338.83.38

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No. PCT/CA99/00852

I. Basis of the report

1. This report has been drawn on the basis of *(substitute sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to the report since they do not contain amendments (Rules 70.16 and 70.17).)*:
Description, pages:

1,3-35	as originally filed	
2,2a	with telefax of	11/12/2000

Claims, No.:

1-14	with telefax of	11/12/2000

2. With regard to the **language**, all the elements marked above were available or furnished to this Authority in the language in which the international application was filed, unless otherwise indicated under this item.

   These elements were available or furnished to this Authority in the following language: , which is:

   - ☐ the language of a translation furnished for the purposes of the international search (under Rule 23.1(b)).
   - ☐ the language of publication of the international application (under Rule 48.3(b)).
   - ☐ the language of a translation furnished for the purposes of international preliminary examination (under Rule 55.2 and/or 55.3).

3. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, the international preliminary examination was carried out on the basis of the sequence listing:

   - ☐ contained in the international application in written form.
   - ☐ filed together with the international application in computer readable form.
   - ☐ furnished subsequently to this Authority in written form.
   - ☐ furnished subsequently to this Authority in computer readable form.
   - ☐ The statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.
   - ☐ The statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished.

4. The amendments have resulted in the cancellation of:

   - ☐ the description, pages:
   - ☐ the claims, Nos.:
   - ☐ the drawings, sheets:

5. ☒ This report has been established as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed (Rule 70.2(c)):

(Any replacement sheet containing such amendments must be referred to under item 1 and annexed to this report.)
see separate sheet

6. Additional observations, if necessary:

III. Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

1. The questions whether the claimed invention appears to be novel, to involve an inventive step (to be non-obvious), or to be industrially applicable have not been examined in respect of:

☐ the entire international application.

☒ claims Nos. 1-8 and 13-14 partially with respect to industrial applicability.

because:

☒ the said international application, or the said claims Nos. 1-8 and 13-14 partially relate to the following subject matter which does not require an international preliminary examination (*specify*):
see separate sheet

☐ the description, claims or drawings (*indicate particular elements below*) or said claims Nos. are so unclear that no meaningful opinion could be formed (*specify*):

☐ the claims, or said claims Nos. are so inadequately supported by the description that no meaningful opinion could be formed.

☐ no international search report has been established for the said claims Nos. .

2. A meaningful international preliminary examination report cannot be carried out due to the failure of the nucleotide and/or amino acid sequence listing to comply with the standard provided for in Annex C of the Administrative Instructions:

☐ the written form has not been furnished or does not comply with the standard.

☐ the computer readable form has not been furnished or does not comply with the standard.

V. Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Statement

Novelty (N)	Yes:	Claims	1-14
	No:	Claims	
Inventive step (IS)	Yes:	Claims	1-14

	No:	Claims	
Industrial applicability (IA)	Yes:	Claims	9-12, 14
	No:	Claims	

2. Citations and explanations
see separate sheet

VIII. Certain observations on the international application

The following observations on the clarity of the claims, description, and drawings or on the question whether the claims are fully supported by the description, are made:
see separate sheet

The arguments filed by the applicant with a letter of 8.12.2000 have been taken into account for establishing said report.

Point I:

The amendments filed with the letter dated 8.12.2000 introduce subject-matter which extends beyond the content of the application as filed, contrary to Article 34(2)(b) PCT. The amendments concerned are the following:
- **claim 8**: non-cancerous cells; no basis for the general term could be found in the originally filed description (p 12, l 19-20 only disclose lymphocytes).
- **claim 9**: "tissue"
- **claim 12**: no basis could be found for the claimed method starting from step b): "assaying a function...as compared to in the absence thereof".
- **page 2a**: "**germline**" mutation; no basis could be found for mutation in this specific cell type. No implicit disclosure is present for the following reasons. A predisposition for breast cancer can be created for instance due to a spontaneous mutation of DNA in other cells than germline cells. The IPEA furthermore believes that such a predisposition can well be diagnosed from somatic cells and does not need to be examined in germline cells only.

Point III:

Claims 1-8, as well as **claim 13** as long as they depend from any one of claims 1-8, relate to subject-matter considered by this Authority to be covered by the provisions of Rule 67.1(iv) PCT. Consequently, no opinion will be formulated with respect to the industrial applicability of the subject-matter of these claims (Article 34(4)(a)(i) PCT).

Point V:

Reference is made to the following documents:

D1: AMERICAN JOURNAL OF HUMAN GENETICS,
vol. 61, no. 4 suppl, 28.10.1997 - 1.11.1997, page A64

D2: WO97/17469

1. Articles 33(2) and (3) PCT

The subject-matter of **claims 1-14** is **novel** (Article 33(2) PCT) in the light of the closest prior art D1, since it is distinguished therefrom in that the androgen receptor is used for determining an individual's predisposition of breast cancer or for screening and selecting an agent which modulates said predisposition.
Furthermore, D1 (l 11 from the bottom) only suggests a correlation of CAG repeat length of AR and breast cancer.

D2 reveals a method of predicting the risk of prostate cancer morbidity and mortality comprising determining the length of the CAG repeat of the androgen receptor gene (abstract, claim 1).
The teaching of said document can not be combined with D1 for the following reasons: a particular marker for a particular cancer (prostate (D1) or breast (D2) cancer) can not be directly transposed to a different type of cancer due to the complexity of the genetic regulation which operates at different hormonal receptors and the intricate interactions of different hormones which can differently affect, in tissue specific fashion, the transactivation of genes they regulate.

Thus, the subject-matter of **claims 1-14** is also **inventive** according to Article 33(3) PCT.

3. Industrial Applicability

For the assessment of the present **claims 1-8**, as well as **claim 13** as long as they depend from any one of claims 1-8, on the question whether they are industrially applicable, no unified criteria exist in the PCT contracting States. The patentability can also be dependent upon the formulation of the claims. The EPO, for example, does not recognize as industrially applicable the subject-matter of claims to the use of a compound in medical treatment, but may allow, however, claims to a known compound for first use in medical treatment and the use of such a compound for the manufacture

of a medicament for a new medical treatment.

Point VIII:

1. The vague and imprecise statement "spirit ... of the subject invention" (page 34, line 12) implies that the subject-matter for which protection is sought may be different to that defined in the claims, thereby resulting in lack of clarity of the claims (Article 6 PCT) when used to interpret them (see the Guidelines, C-III, 4.3a). The above statement has not been deleted to remove this defect.

2. The wording of claim 12 is unclear (Article 6 PCT), since it cannot be derived which kind of function ("**a** function") of the allele is assayed.

3. The broad terms "variant", "equivalent" and "mutation" used in claims 1 and 12 are vague and unclear and leave the reader in doubt as to the meaning of the technical features to which they refer, thereby rendering the definition of the subject-matter of said claim unclear (Article 6 PCT).
Furthermore, the definitions given in the description (page 21, lines 1-17, lines 18-20 and page 22, lines 3-10) do not clarify the terms, since it is not said to what extent these substances are allowed to differ from the androgen receptor gene. The terms "variant" and "mutation" are even so broad that they encompass any nucleic acid molecule, since they are not even limited by the function/s of the androgen receptor gene. Moreover, the definition of the term "equivalent" is as vague since it is silent about how many and which function/s have to remain compared to the androgen receptor gene.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12Q 1/68

A2

(11) International Publication Number: WO 00/15834

(43) International Publication Date: 23 March 2000 (23.03.00)

(21) International Application Number: PCT/CA99/00852

(22) International Filing Date: 15 September 1999 (15.09.99)

(30) Priority Data:
60/100,311 15 September 1998 (15.09.98) US

(71) Applicant *(for all designated States except US)*: SIGNALGENE INC. [CA/CA]; 8475 Christophe Colomb Avenue, Montreal, Quebec H2M 2N9 (CA).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: ROUSSEAU, François [CA/CA]; 2540 Lalonde, Sainte-Foy, Quebec G1W 1M7 (CA).

(74) Agents: DUBUC, Jean, H. et al.; Goudreau Gage Dubuc & Martineau Walker, The Stock Exchange Tower, Suite 3400, 800 Place Victoria, P.O. Box 242, Montreal, Quebec H4Z 1E9 (CA).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: MARKER AT THE ANDROGEN RECEPTOR GENE FOR DETERMINING BREAST CANCER SUSCEPTIBILITY

(57) Abstract

The present invention relates to a method of determining an individual's predisposition to breast cancer, development of breast cancer, protection against breast cancer and/or responsiveness to therapy for breast cancer. The method comprises the step of determining the androgen receptor genotype at the CAG repeat locus of an individual, or at a locus in linkage disequilibrium with the CAG repeat locus, thereby determining an individual's predisposition to breast cancer, development of breast cancer, protection against breast cancer and/or responsiveness to therapy for breast cancer.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

Breast cancers have the same clinical characteristics in older as in younger women. Cancer is usually suspected when changes are noted on mammography or when a breast lesion is seen or felt. Lesions usually can be felt as firm nodules within the breast. Ulcerations may occur, and lesions within or near the nipple may produce discharge. Sometimes breast cancer is discovered only after metastatic lesions cause bone fractures, neurologic changes, hypercalcemia, liver failure, or ascites.

When a tumor is detected by physical examination, bilateral mammograms are normally obtained to rule out occult lesions. Certain radiographic images, such as speckled calcifications or tissue infiltration, suggest cancer, while a cystic appearance suggests a benign process. Even an apparently benign finding on mammogram requires further evaluation. Generally the diagnosis is established by fine needle aspiration. Fine needle aspiration allows collection and cytological examination of cystic fluid and is helpful in planning definitive treatment of breast cancer. Although a positive result on fine needle aspiration is diagnostic, a negative result is usually followed by an open biopsy. Now a day, there is still no specific test for assaying predisposition or resistance to breast cancer.

Since the discovery of the human androgen receptor (AR) gene, mutations in this gene have been associated with Kennedy's disease (spinobulbarmuscolar atrophy), with various degrees of androgen insensitivity and with prostate cancer. Thus, an association between the AR gene and breast cancer has yet to be reported.

There thus remains a need to provide a genetic assay for determining the predisposition and/or resistance to breast cancer, development of breast cancer and responsiveness to therapeutic modalities.

While some markers have been identified as genetic determinants for breast cancer and/or as risk factors to develop same (i.e. BRCA1 and BRCA2), there remains a need to identify new markers therefor. More specifically, there remains a need to provide means to determine a

predisposition to breast cancer and/or responsiveness to therapy to breast cancer, by analyzing allelic variations in genes associated with breast cancer. In addition there remains a need to identify patients who are likely to benefit from a particular prevention or therapeutic treatment program. Further, there remains a need to provide assays to screen for compounds (i.e. hormones, molecules acting on hormone receptors or other agents) that could be beneficial to patients.

The present invention seeks to meet these and other needs.

The present description refers to a number of documents, the content of which is herein incorporated by reference, in their entirety.

SUMMARY OF THE INVENTION

One aim of the present invention is to provide a genetic assay for determining the predisposition to breast cancer and/or response to breast cancer treatment.

Another aim of the present invention is to use a polymorphism of the androgen receptor (AR) gene or an equivalent thereof as a marker for breast cancer susceptibility and/or response to breast cancer preventive or curative therapy. A polymorphism of the androgen receptor (AR) gene, or any polymorphism in linkage disequilibrium therewith, can be used as a test for breast cancer susceptibility, for responsiveness to treatment of breast cancer, for breast cancer prognosis or severity, or as a means to classify patients in clinical trial for breast cancer (screening, diagnosis, prognosis or treatment).

One of a polymorphism of the AR gene, or any polymorphism in linkage disequilibrium therewith, can further be used as a test for screening drugs for breast cancer or for determining the best treatment therefor.

Broadly, the present invention aims at providing a method of determining the length of a CAG repeat polymorphism in the androgen receptor gene, wherein this determination can be correlated with a predisposition or a

protection to breast cancer. This determination can be based on a variety of genotyping methods at the DNA, RNA or protein level.

Another aim of the present invention is to provide a method of prognosing and/or forecasting the development of breast cancer in a patient, which comprises determining a CAG-repeat polymorphism of the AR gene, or any polymorphism in linkage disequilibrium therewith, in a biological sample of the patient, wherein a determination of the length of the CAG repeat shows a significant association with breast cancer.

In a particular embodiment, the determination of the polymorphism at the CAG repeat of the AR gene enables to show that the shortest alleles or a combination of the shortest alleles are associated with the smallest breast cancer risk and the mid to long alleles or a combination of the intermediate and longest alleles are associated with the highest breast cancer risk (a combination of the longest alleles is associated with the highest risks of breast cancer) . Of importance, the variations of polymorphisms at the CAG repeat locus of AR (or of an equivalent or marker in linkage disequilibrium therewith) can account for a significant proportion of all cases of breast cancer. Indeed, the number of breast cancer cases attributable to a variation at this AR locus is at least three times greater than that attributable to the BRCA1 and BRCA2 genes.

The present invention also relates to vectors, including expression vectors harboring an AR gene (or fragment or fusion thereof) having a genotype in accordance with the present invention (i.e. a predisposing genotype, long CAG repeats, or alternatively, a protecting genotype, short CAG repeats; or other genotypes isolated from patients or genetically engineered), cells harboring such vectors, and non-human animals harboring such vectors or cells.

Another aim of the present invention is to provide means of identifying young women that will be at risk of developing breast cancer and to categorize those that are likely to respond significantly to preventive therapy.

An aim of the present invention is thus to provide means of identification of target sub-groups of women for breast cancer prevention measures/programs.

Another aim of the present invention is to provide means to determine which sub-group of women will most benefit from breast cancer treatment(s) and eventually predict their response to therapy or choose the optimal preventive pharmacotherapy.

Another aim of the present invention is to identify means of predicting and managing interventions for breast cancer as well as identifying and/or characterizing biological parameters which could enable the establishment of population-based breast cancer prevention and intervention programs.

In addition, it is an aim of the present invention to provide a method of selecting alleles of the AR gene or in linkage disequilibrium therewith, which is suitable for designing an assay to screen compounds which can modulate the activity of an androgen receptor.

Another aim of the present invention is to provide an assay to screen for drugs for the treatment and/or prevention of breast cancer. Having identified alleles which predispose to breast cancer (and those which predispose to a "resistance" to breast cancer), assays can be set-up to screen agents and select drugs which could be used in the treatment or prevention of breast cancer. Since some alleles of the AR have been shown to affect the functionality of the androgen receptor (Tut et al. 1997, J. Clin. Endocrinol. 89(11):3777-3782), assays could be designed based on chosen genotypes of the AR gene. A non-limiting example of a type of assay which could be designed includes, *cis-trans* assays similar to those described in USP 4,981,784. For example, a *cis-trans* assay could be set-up, based on the use of a genotype of AR, shown here to predispose to breast cancer (i.e. the long CAG alleles in the AR gene) as compared to a genotype of AR, shown here to be associated with lower risk of breast cancer, and used to screen compounds. A non-limiting example of such an assay could be based on 2 cell lines (one

expressing a predisposing genotype of AR and one expressing a non-predisposing genotype of AR) which could be used in parallel to screen for AR-function modulating compounds. Of course, it will be understood that the cell line expressing the non-predisposing genotype of AR (the shorter alleles) can be used as a positive control for the functionality of the androgen receptor.

It is thus an aim of the present invention to provide the means to identify compounds which could positively modulate the function of AR having a breast cancer predisposing genotype (such as the long CAG alleles), to the level of the protecting genotype thereof (such as the short CAG alleles).

In a particular embodiment, such assays can be designed using cells from patients having a known genotype at the loci of the present invention, these cells harboring recombinant vectors could enable an assessment of the functionality of the AR and dissect the structure-function relationship of the androgen receptor and its role in breast cancer.

It shall be understood that the polymorphism of the AR and/or the determination of allelic variations in the AR gene can be combined to the determination of allelic variations in other genes/markers linked to the predisposition to breast cancer and/or responsiveness to therapy therefor. This combination of genotype analyses could lead to better diagnoses programs and/or treatment of breast cancer. Non-limiting examples of such markers include BRCA1 and BRCA2.

It shall also be understood that although breast cancer is significantly more preponderant in women, it can also be a deadly disease in men. Thus, the present invention is meant to also cover men.

In accordance with the present invention, there is therefore provided a method of determining an individual's predisposition to breast cancer, development of breast cancer and/or responsiveness to therapy for breast cancer, which comprises determining a genotype at the CAG-repeat locus of the androgen receptor (directly or indirectly by linkage disequilibrium) in a biological sample of the individual and analyzing allelic variation in the androgen receptor

of the individual, thereby determining an individual's predisposition to breast cancer, development of breast cancer and/or responsiveness to therapy therefor.

In accordance with the present invention there is provided a method for determining susceptibility to breast cancer, and/or response to therapy therefor. The method comprises the step of determining the androgen receptor genotype of the individual, thereby determining an individual's susceptibility to breast cancer and/or response to therapy therefor.

Numerous methods for determining a genotype are known and availble to the skilled artisan. All these genotype determination methods are within the scope of the present invention. Non-limiting examples of genotype determination include a restriction endonuclease digestion, a hybridization with allele specific oligonucleotides, a sequencing of the polymorphism, and an amplification of a segment of the androgen receptor (i.e. by PCR).

In accordance with the present invention, there is therefore provided a method of determining an individual's predisposition to breast cancer, development of breast cancer and/or responsiveness to therapy therefor, which comprises determining androgen receptor polymorphism (directly or indirectly using a marker in linkage disequilibrium with the CAG repeat polymorphism) in a biological sample of the individual and analyzing allelic variation in the androgen receptor gene of the individual, thereby determining an individual's predisposition to breast cancer, development of breast cancer and/or responsiveness to therapy therefor.

In accordance with one embodiment of the invention, there is provided a specific model for use in prediction of breast cancer susceptibility and prognosis. The model comprises an androgen receptor gene polymorphisms at the CAG repeat locus, that allows to identify a subset of women that are at significantly increased risk of breast cancer as compared to those bearing other variant of this gene.

In accordance with a preferred embodiment of the present invention, a single gene, the androgen receptor gene, has been identified as such a target to assess this predisposition.

In accordance with the present invention, the androgen receptor polymorphism, without limitation, is selected from the CAG repeats located in the first exon of the AR gene, or any DNA variant or mutation which shows some degree of linkage disequelibrium with one of the polymorphisms at the CAG-repeat locus of the AR gene.

In some embodiments, the method of the present invention includes detecting the androgen receptor polymorphism by analyzing the restriction fragment length polymorphisms using an endonuclease digestion. The method can further include a step prior to the androgen receptor gene digestion, wherein at least a fragment of the androgen receptor is amplified, for example, by polymerase chain reaction.

In accordance with a preferred embodiment of the present invention, a pair of primers is designed to specifically amplify a segment of the androgen receptor. In an especially preferred embodiment, the region of the AR gene which is amplified is in exon 1. This pair of primers is preferably derived from a nucleic acid sequence of the androgen receptor gene or flanking portion thereof, to amplify a segment of the androgen receptor gene, as commonly known. Of course, other primer pairs can be designed, based on the known sequence of the AR gene. Method to design primer pairs form known sequences are commonly known in the art.

In accordance with a preferred embodiment of the present invention, primers used for amplifying the segment of the androgen receptor are defined as:

5'-TCCAGAATCTGTTCCAGAGCGTGC-3 (SEQ ID NO:1); and

5'-GCTGTGAAGGTTGCTGTTCCTCAT-3' (SEQ ID NO:2).

The polymorphism of the androgen receptor gene can be detected using at least one oligonucleotide specific to the normal or variant androgen receptor gene allele.

The present invention also provides a kit for determining predisposition to low, intermediate or high risk of breast cancer of a patient, which includes at least a probe specific for the androgen receptor; a polymorphism selected from the group consisting of a CAG repeat and other polymorphisms in linkage disequilibrium with the CAG repeat polymorphism.

In one embodiment, the present invention provides a specific detection of the CAG repeat polymorphism of the AR gene using a nucleic acid for the specific detection of this AR polymorphism in a sample comprising the above-described CAG-repeat-containing nucleic acid sequence (i.e. DNA, RNA, cDNA) and at least a nucleic acid sequence which binds under stringent conditions to the CAG-repeat-containing nucleic acid sequence.

In one prefered embodiment, the present invention relates to nucleic acid probes which are complementary to a CAG-repeat-containing nucleic acid sequence, consisting of at least 10 consecutive nucleotides (preferably, 15, 20, 25, or 30) and which specifically hybridize to the AR nucleic acid sequence comprising the CAG repeat polymorphism under high stringency condition.

In one embodiment of the above described method, a nucleic acid probe is immobilized on a solid support. Non-limiting examples of solid supports include plastics (i.e.polycarbonate), acrylic resins (i.e. polyacrylamide and latex beads); and carbohydrates (i.e. agarose and sepharose). Techniques for coupling nucleic acid probes to solid supports are well known in the art.

Similarly to the probes of the present invention, the antibodies of the present invention can be immobilized on a solid support. As known in the art, similar supports as those used for probe immobilization can be used for antibody immobilization on a solid support. Also well known in the art are the techniques for coupling antibodies to such solid supports. The immobilized

antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays as well as in immunochromatography according to known methods.

Non-limiting examples of test samples suitable for carrying the methods of the present invention include, cells or nucleic acid extracts of cells, or biological fluids. Of course, the type of test sample used can vary according to the assay format, the method of detection, and the particular needs of the clinical practioner which will readily adapt the methods of preparation of the sample and the method of detection so that they are compatible, in accordance with the knowledge in the art.

In accordance with one embodiment of the present invention, the allelic variation in the androgen receptor gene is analyzed indirectly using a nucleic acid variant, or equivalent in linkage disequilibrium with a CAG repeat. The allelic variation in the androgen receptor gene can also be analyzed directly by determining the number of CAG repeat within the androgen receptor gene.

In accordance with the present invention, the polymorphism of the androgen receptor (AR) gene can be used as a marker for breast cancer susceptibility. The polymorphism in linkage disequilibrium with the markers used can also be used as a test for breast cancer susceptibility, or for responsiveness to treatment for breast cancer, for breast cancer prognosis or severity, or as a means to classify patients in clinical trials for breast cancer (screening, diagnosis, prognosis or treatment).

In order to provide a clear and consistent understanding of terms used in the present description, a number of definitions are provided hereinbelow.

As used herein the term "RFLP" refers to restriction fragment length polymorphism.

The terms "polymorphism", "DNA polymorphism" and the like, refer to any sequence in the human genome which exists in more than one version or variant in the population.

The term "linkage disequilibrium" refers to any degree of non-random genetic association between one or more allele(s) of two different polymorphic DNA sequences, that is due to the physical proximity of the two loci. Linkage disequilibrium is present when two DNA segments that are very close to each other on a given chromosome will tend to remain unseparated for several generations with the consequence that alleles of a DNA polymorphism (or marker) in one segment will show a non-random association with the alleles of a different DNA polymorphism (or marker) located in the other DNA segment nearby. Hence, testing of one of a marker in linkage desiquilibrium with the polymorphisms of the present invention at the AR gene (indirect testing), will give almost the same information as testing for the CAG repeat polymorphism of the AR gene directly. This situation is encountered throughout all the human genome when two DNA polymorphisms that are very close to each other are studied. Such a linkage disequilibrium has been reported with several polymorphisms in several genes (i.e. the vitamin D receptor gene [Morrisson et al., 1994, Nature 367:284-287, and USP 5,593,033]). Various degrees of linkage disequilibrium can be encountered between two genetic markers so that some are more closely associated than others.

The terms "androgen receptor polymorphism" or "genetic marker" are intended to include, without limitation, the CAG-repeat polymorphism in exon 1, and any other allelic variant of the androgen receptor gene that show some degree of linkage disequilibrium in any population sub-group with at least one of the above-mentioned androgen receptor polymorphisms.

The androgen receptor gene polymorphism sites in accordance with the present invention can be located within the androgen receptor gene, or on each side thereof, provided that is on the same chromosome and in linkage disequilibrium with the AR polymorphism of the present invention. Distances between markers in linkage disequilibrium can vary widely (below 50 kb to more than 1 mega base) depending on the genetic

structure of the population and is ascertainable by a statistically significant association between the markers.

It shall be recognized by the person skilled in the art to which the present invention pertains, that since some of the polymorphisms herein identified in the AR gene can be within the coding region of the gene and therefore expressed, that the present invention should not be limited to the identification of polymorphisms at the DNA level (whether on genomic DNA, amplified DNA, cDNA or the like). Indeed, the herein-identified polymorphisms could be detected at the mRNA or protein level. Such detections of polymorphism identification on mRNA or protein are known in the art. Non-limiting examples include detection based on oligos designed to hybridize to mRNA or ligands such as antibodies which are specific to the encoded polymorphism (i.e. specific to the protein fragment encoded by the CAG repeat for example).

Since some of the polymorphisms of the present invention are expressed, one of the advantages of the present invention is to enable a determination of the polymorphisms in the AR gene, in easily obtainable cells which express these genes. A non-limiting example thereof is lymphocytes, thereby enabling a genotyping from a simple blood sample.

Nucleotide sequences are presented herein by single strand, in the 5' to 3' direction, from left to right, using the one letter nucleotide symbols as commonly used in the art and in accordance with the recommendations of the IUPAC-IUB Biochemical Nomenclature Commission.

Unless defined otherwise, the scientific and technological terms and nomenclature used herein have the same meaning as commonly understood by a person of ordinary skill to which this invention pertains. Generally, the procedures for cell cultures, infection, molecular biology methods and the like are common methods used in the art. Such standard techniques can be found in reference manuals such as for example Sambrook et al. (1989,

Molecular Cloning -A Laboratory Manual, Cold Spring Harbor Laboratories) and Ausubel et al. (1994, Current Protocols in Molecular Biology, Wiley, New York).

The present description refers to a number of routinely used recombinant DNA (rDNA) technology terms. Nevertheless, definitions of selected examples of such rDNA terms are provided for clarity and consistency.

As used herein, "nucleic acid molecule", refers to a polymer of nucleotides. Non-limiting examples thereof include DNA (i.e. genomic DNA, cDNA) and RNA molecules (i.e. mRNA). The nucleic acid molecule can be obtained by cloning techniques or synthesized. DNA can be double-stranded or single-stranded (coding strand or non-coding strand [antisense]).

The term "recombinant DNA" as known in the art refers to a DNA molecule resulting from the joining of DNA segments. This is often referred to as genetic engineering.

The term "DNA segment", is used herein, to refer to a DNA molecule comprising a linear stretch or sequence of nucleotides. This sequence when read in accordance with the genetic code, can encode a linear stretch or sequence of amino acids which can be referred to as a polypeptide, protein, protein fragment and the like.

The terminology "amplification pair" refers herein to a pair of oligonucleotides (oligos) of the present invention, which are selected to be used together in amplifying a selected nucleic acid sequence by one of a number of types of amplification processes, preferably a polymerase chain reaction. Other types of amplification processes include ligase chain reaction, strand displacement amplification, or nucleic acid sequence-based amplification, as explained in greater detail below. As commonly known in the art, the oligos are designed to bind to a complementary sequence under selected conditions.

The nucleic acid (i.e. DNA or RNA) for practicing the present invention may be obtained according to well known methods.

Oligonucleotide probes or primers of the present invention may be of any suitable length, depending on the particular assay format and the

particular needs and targeted genomes employed. In general, the oligonucleotide probes or primers are at least 12 nucleotides in length, preferably between 15 and 24 molecules, and they may be adapted to be especially suited to a chosen nucleic acid amplification system. As commonly known in the art, the oligonucleotide probes and primers can be designed by taking into consideration the melting point of hydrizidation thereof with its targeted sequence (see below and in Sambrook et al., 1989, Molecular Cloning -A Laboratory Manual, 2nd Edition, CSH Laboratories; Ausubel et al., 1989, in Current Protocols in Molecular Biology, John Wiley & Sons Inc., N.Y.).

The term "oligonucleotide" or "DNA" molecule or sequence refers to a molecule comprised of the deoxyribonucleotides adenine (A), guanine (G), thymine (T) and/or cytosine (C), in a double-stranded form, and comprises or includes a "regulatory element" according to the present invention, as the term is defined herein. The term "oligonucleotide" or "DNA" can be found in linear DNA molecules or fragments, viruses, plasmids, vectors, chromosomes or synthetically derived DNA. As used herein, particular double-stranded DNA sequences may be described according to the normal convention of giving only the sequence in the 5' to 3' direction.

"Nucleic acid hybridization" refers generally to the hybridization of two single-stranded nucleic acid molecules having complementary base sequences, which under appropriate conditions will form a thermodynamically favored double-stranded structure. Examples of hybridization conditions can be found in the two laboratory manuals referred above (Sambrook et al., 1989, *supra* and Ausubel et al., 1989, *supra*) and are commonly known in the art. In the case of a hybridization to a nitrocellulose filter, as for example in the well known Southern blotting procedure, a nitrocellulose filter can be incubated overnight at 65°C with a labeled probe in a solution containing 50% formamide, high salt (5 x SSC or 5 x SSPE), 5 x Denhardt's solution, 1% SDS, and 100 µg/ml denatured carrier DNA (i.e. salmon sperm DNA). The non-specifically binding probe can then be washed off the

filter by several washes in 0.2 x SSC/0.1% SDS at a temperature which is selected in view of the desired stringency: room temperature (low stringency), 42°C (moderate stringency) or 65°C (high stringency). The selected temperature is based on the melting temperature (Tm) of the DNA hybrid. Of course, RNA-DNA hybrids can also be formed and detected. In such cases, the conditions of hybridization and washing can be adapted according to well known methods by the person of ordinary skill. Stringent conditions will be preferably used (Sambrook et al.,1989, *supra*).

Probes of the invention can be utilized with naturally occurring sugar-phosphate backbones as well as modified backbones including phosphorothioates, dithionates, alkyl phosphonates and α-nucleotides and the like. Modified sugar-phosphate backbones are generally taught by Miller, 1988, Ann. Reports Med. Chem. 23:295 and Moran et al., 1987, Nucleic acid molecule. Acids Res., 14:5019. Probes of the invention can be constructed of either ribonucleic acid (RNA) or deoxyribonucleic acid (DNA), and preferably of DNA.

The types of detection methods in which probes can be used include Southern blots (DNA detection), dot or slot blots (DNA, RNA), and Northern blots (RNA detection). Although less preferred, labeled proteins could also be used to detect a particular nucleic acid sequence to which it binds. More recently, PNAs have been described (Nielsen et al. 1999, Current Opin. Biotechnol. 10:71-75). PNAs could also be used to detect the polymorphisms of the present invention. Other detection methods include kits containing probes on a dipstick setup and the like.

Although the present invention is not specifically dependent on the use of a label for the detection of a particular nucleic acid sequence, such a label might be beneficial, by increasing the sensitivity of the detection. Furthermore, it enables automation. Probes can be labeled according to numerous well known methods (Sambrook et al., 1989, supra). Non-limiting examples of labels include ^{3}H, ^{14}C, ^{32}P, and ^{35}S. Non-limiting examples of detectable markers include ligands, fluorophores, chemiluminescent agents,

enzymes, and antibodies. Other detectable markers for use with probes, which can enable an increase in sensitivity of the method of the invention, include biotin and radionucleotides. It will become evident to the person of ordinary skill that the choice of a particular label dictates the manner in which it is bound to the probe.

As commonly known, radioactive nucleotides can be incorporated into probes of the invention by several methods. Non-limiting examples thereof include kinasing the 5' ends of the probes using gamma ^{32}P ATP and polynucleotide kinase, using the Klenow fragment of Pol I of *E. coli* in the presence of radioactive dNTP (i.e. uniformly labeled DNA probe using random oligonucleotide primers in low-melt gels), using the SP6/T7 system to transcribe a DNA segment in the presence of one or more radioactive NTP, and the like.

As used herein, "oligonucleotides" or "oligos" define a molecule having two or more nucleotides (ribo or deoxyribonucleotides). The size of the oligo will be dictated by the particular situation and ultimately on the particular use thereof and adapted accordingly by the person of ordinary skill. An oligonucleotide can be synthesized chemically or derived by cloning according to well known methods.

As used herein, a "primer" defines an oligonucleotide which is capable of annealing to a target sequence, thereby creating a double stranded region which can serve as an initiation point for DNA synthesis under suitable conditions.

Amplification of a selected, or target, nucleic acid sequence may be carried out by a number of suitable methods. See generally Kwoh et al., 1990, Am. Biotechnol. Lab. 8:14-25. Numerous amplification techniques have been described and can be readily adapted to suit particular needs of a person of ordinary skill. Non-limiting examples of amplification techniques include polymerase chain reaction (PCR), ligase chain reaction (LCR), strand displacement amplification (SDA), transcription-based

amplification, the Qβ replicase system and NASBA (Kwoh et al., 1989, Proc. Natl. Acad. Sci. USA 86, 1173-1177; Lizardi et al., 1988, BioTechnology 6:1197-1202; Malek et al., 1994, Methods Mol. Biol., 28:253-260; and Sambrook et al., 1989, *supra*). Preferably, amplification will be carried out using PCR.

Polymerase chain reaction (PCR) is carried out in accordance with known techniques. See, e.g., U.S. Pat. Nos. 4,683,195; 4,683,202; 4,800,159; and 4,965,188 (the disclosures of all three U.S. Patent are incorporated herein by reference). In general, PCR involves, a treatment of a nucleic acid sample (e.g., in the presence of a heat stable DNA polymerase) under hybridizing conditions, with one oligonucleotide primer for each strand of the specific sequence to be detected. An extension product of each primer which is synthesized is complementary to each of the two nucleic acid strands, with the primers sufficiently complementary to each strand of the specific sequence to hybridize therewith. The extension product synthesized from each primer can also serve as a template for further synthesis of extension products using the same primers. Following a sufficient number of rounds of synthesis of extension products, the sample is analysed to assess whether the sequence or sequences to be detected are present. Detection of the amplified sequence may be carried out by visualization following EtBr staining of the DNA following gel electrophores, or using a detectable label in accordance with known techniques, and the like. For a review on PCR techniques (see PCR Protocols, A Guide to Methods and Amplifications, Michael et al. Eds, Acad. Press, 1990).

Ligase chain reaction (LCR) is carried out in accordance with known techniques (Weiss, 1991, Science 254:1292). Adaptation of the protocol to meet the desired needs can be carried out by a person of ordinary skill. Strand displacement amplification (SDA) is also carried out in accordance with known techniques or adaptations thereof to meet the particular needs (Walker et al., 1992, Proc. Natl. Acad. Sci. USA 89:392-396; and ibid., 1992, Nucleic Acids Res. 20:1691-1696).

As used herein, the term "gene" is well known in the art and relates to a nucleic acid sequence defining a single protein or polypeptide. A "structural gene" defines a DNA sequence which is transcribed into RNA and translated into a protein having a specific amino acid sequence thereby giving rise the a specific polypeptide or protein.

A "heterologous" (i.e. a heterologous gene) region of a DNA molecule is a subsegment segment of DNA within a larger segment that is not found in association therewith in nature. The term "heterologous" can be similarly used to define two polypeptidic segments not joined together in nature. Non-limiting examples of heterologous genes include reporter genes such as luciferase, chloramphenicol acetyl transferase, β-galactosidase, and the like which can be juxtaposed or joined to heterologous control regions or to heterologous polypeptides.

The term "vector" is commonly known in the art and defines a plasmid DNA, phage DNA, viral DNA and the like, which can serve as a DNA vehicle into which DNA of the present invention can be cloned. Numerous types of vectors exist and are well known in the art.

The term "expression" defines the process by which a gene is transcribed into mRNA (transcription), the mRNA is then being translated (translation) into one polypeptide (or protein) or more.

The terminology "expression vector" defines a vector or vehicle as described above but designed to enable the expression of an inserted sequence following transformation into a host. The cloned gene (inserted sequence) is usually placed under the control of control element sequences such as promoter sequences. The placing of a cloned gene under such control sequences is often refered to as being operably linked to control elements or sequences.

Operably linked sequences may also include two segments that are transcribed onto the same RNA transcript. Thus, two sequences, such as a promoter and a "reporter sequence" are operably linked

if transcription commencing in the promoter will produce an RNA transcript of the reporter sequence. In order to be "operably linked" it is not necessary that two sequences be immediately adjacent to one another.

Expression control sequences will vary depending on whether the vector is designed to express the operably linked gene in a prokaryotic or eukaryotic host or both (shuttle vectors) and can additionally contain transcriptional elements such as enhancer elements, termination sequences, tissue-specificity elements, and/or translational initiation and termination sites.

Prokaryotic expressions are useful for the preparation of large quantities of the protein encoded by the DNA sequence of interest. This protein can be purified according to standard protocols that take advantage of the intrinsic properties thereof, such as size and charge (i.e. SDS gel electrophoresis, gel filtration, centrifugation, ion exchange chromatography...). In addition, the protein of interest can be purified via affinity chromatography using polyclonal or monoclonal antibodies. The purified protein can be used for therapeutic applications.

The DNA construct can be a vector comprising a promoter that is operably linked to an oligonucleotide sequence of the present invention, which is in turn, operably linked to a heterologous gene, such as the gene for the luciferase reporter molecule. "Promoter" refers to a DNA regulatory region capable of binding directly or indirectly to RNA polymerase in a cell and initiating transcription of a downstream (3' direction) coding sequence. For purposes of the present invention, the promoter is bound at its 3' terminus by the transcription initiation site and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter will be found a transcription initiation site (conveniently defined by mapping with S1 nuclease), as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase. Eukaryotic promoters will often, but not always,

contain "TATA" boses and "CCAT" boxes. Prokaryotic promoters contain Shine-Dalgarno sequences in addition to the -10 and -35 consensus sequences.

In accordance with one embodiment of the present invention, an expression vector can be constructed to assess the functionality of specific alleles of the AR gene and of the interaction of such alleles. Non-limiting examples of such expression vectors include a vector comprising the androgen responsive element (the cis sequences [i.e. DNA sequence to which a factor binds] enabling androgen-dependent modulating effects of promoter activity are known in the art) operably linked to a chosen promoter and modulating the activity thereof, the promoter driving the expression of a reporter gene. When such a vector is tranfected in a cell expressing AR, the modulating effect of the promoter activity can be assessed by determining the level of expression of the reporter gene. In one embodiment, the vector is transfected into a cell of a patient having the genotype of AR shown herein to be associated with a low risk of breast cancer, or in a cell from a patient having the genotype of AR shown herein to be associated with a moderate or high risk of breast cancer. These cells can serve to screen for compounds that modulate the promoter activity, in order to identify compounds that could be used to treat especially, patients predicted to be at moderate or high risk of breast cancer. Of course, it will be understood that the AR gene expressed by these cells can be modified at will (i.e. by *in vitro* mutagenesis or the like). Similarly, numerous combinations of genotypes can be tested in such assays to dissect the functional relationship between the AR genotype and its function in androgen-dependent function and/or its function in breast cancer. It will also be clear to the skilled artisan, that such indicator cells expressing AR, could also be engineered by choosing a cell line and transfecting thereinto, chosen genotypes of AR and one expression vector as described above. Non-human transgenic animals expressing chosen alleles of AR could also be prepared and used to screen compounds that affect androgen receptor function and possibly overcome a predisposition to breast cancer.

As used herein, the designation "functional derivative" denotes, in the context of a functional derivative of a sequence whether an nucleic acid or amino acid sequence, a molecule that retains a biological activity (either function or structural) that is substantially similar to that of the original sequence. This functional derivative or equivalent may be a natural derivative or may be prepared synthetically. Such derivatives include amino acid sequences having substitutions, deletions, or additions of one or more amino acids, provided that the biological activity of the protein is conserved. The same applies to derivatives of nucleic acid sequences which can have substitutions, deletions, or additions of one or more nucleotides, provided that the biological activity of the sequence is generally maintained. When relating to a protein sequence, the substituting amino acid as chemico-physical properties which are similar to that of the substituted amino acid. The similar chemico-physical properties include, similarities in charge, bulkiness, hydrophobicity, hydrophylicity and the like. The term "functional derivatives" is intended to include "fragments", "segments", "variants", "analogs" or "chemical derivatives" of the subject matter of the present invention.

Thus, the term "variant" refers herein to a protein or nucleic acid molecule which is substantially similar in structure and biological activity to the protein or nucleic acid of the present invention.

The functional derivatives of the present invention can be synthesized chemically or produced through recombinant DNA technology. all these methods are well known in the art.

As used herein, "chemical derivatives" is meant to cover additional chemical moieties not normally part of the subject matter of the invention. Such moieties could affect the physico-chemical characteristic of the derivative (i.e. solubility, absorption, half life and the like, decrease of toxicity). Such moieties are examplified in Remington's Pharmaceutical Sciences (1980). Methods of coupling these chemical-physical moieties to a polypeptide are well known in the art.

The term "allele" defines an alternative form of a gene which occupies a given locus on a chromosome.

As commonly known, a "mutation" is a detectable change in the genetic material which can be transmitted to a daughter cell. As well known, a mutation can be, for example, a detectable change in one or more deoxyribonucleotide. For example, nucleotides can be added, deleted, substituted for, inverted, or transposed to a new position. Spontaneous mutations and experimentally induced mutations exist. The result of a mutations of nucleic acid molecule is a mutant nucleic acid molecule. A mutant polypeptide can be encoded from this mutant nucleic acid molecule.

As used herein, the term "purified" refers to a molecule having been separated from a cellular component. Thus, for example, a "purified protein" has been purified to a level not found in nature. A "substantially pure" molecule is a molecule that is lacking in all other cellular components.

As used herein, the terms "molecule", "compound", or "agent" are used interchangeably and broadly to refer to natural, synthetic or semi-synthetic molecules or compounds. The term "molecule" therefore denotes for example chemicals, macromolecules, cell or tissue extracts (from plants or animals) and the like. Non limiting examples of molecules include nucleic acid molecules, peptides, ligands, including antibodies, carbohydrates and pharmaceutical agents. The agents can be selected and screened by a variety of means including random screening, rational selection and by rational design using for example protein or ligand modelling methods such as computer modelling. The terms "rationally selected" or "rationally designed" are meant to define compounds which have been chosen based on the configuration of the interaction domains of the present invention. As will be understood by the person of ordinary skill, macromolecules having non-naturally occurring modifications are also within the scope of the term "molecule". For example, peptidomimetics, well known in the pharmaceutical industry and generally referred to as peptide analogs can be generated by modelling as mentioned

above. Similarly, in a preferred embodiment, the polypeptides of the present invention are modified to enhance their stability. It should be understood that in most cases this modification should not alter the biological activity of the protein. The molecules identified in accordance with the teachings of the present invention have a therapeutic value in diseases or conditions in which a apparently lower activity and/or level of the AR is linked to a genotype of AR identified in accordance with the present invention. Alternatively, the molecules identified in accordance with the teachings of the present invention find utility in the development of compounds which can modulate the activity and/or level of the androgen receptor in an animal and/or overcome a predisposition to breast cancer.

As used herein, agonists and antagonists also include potentiators of known compounds with such agonist or antagonist properties. In one embodiment, modulators of the level or the activity of the AR can be identified and selected by contacting the indicator cell with a compound or mixture or library of molecules for a fixed period of time. In certain embodiments, the "breast cancer-low risk-associated alleles"of the AR gene can be used as positive controls.

An indicator cell in accordance with the present invention can be used to identify antagonists. For example, the test molecule or molecules are incubated with the host cell in conjunction with one or more agonists held at a fixed concentration. An indication and relative strength of the antagonistic properties of the molecule(s) can be provided by comparing the level of gene expression in the indicator cell in the presence of the agonist, in the absence of test molecules vs in the presence thereof. Of course, the antagonistic effect of a molecule can also be determined in the absence of agonist, simply by comparing the level of expression of the reporter gene product in the presence and absence of the test molecule(s).

It shall be understood that the *"in vivo"* experimental model can also be used to carry out an *"in vitro"* assay. For example, cellular

extracts from the indicator cells can be prepared and used in an *"in vitro"* test. A non-limiting example thereof include binding assays.

As used herein the recitation "indicator cells" refers to cells that express a given genotype of AR according to the present invention. As alluded to above, such indicator cells can be used in the screening assays of the present invention. In certain embodiments, the indicator cells have been engineered so as to express a chosen derivative, fragment, homolog, or mutant of a genotype of the present invention. The cells can be yeast cells or higher eukaryotic cells such as mammalian cells. In one particular embodiment, the indicator cell would be a yeast cell harboring vectors enabling the use of the two hybrid system technology, as well known in the art (Ausubel et al., 1994, *supra*) and can be used to test a compound or a library thereof. In another embodiment, the *cis-trans* assay as described in USP 4,981,784, can be adapted and used in accordance with the present invention. Such an indicator cell could be used to rapidly screen at high-throughput a vast array of test molecules. In a particular embodiment, the reporter gene is luciferase or β-Gal.

In some embodiments, it might be beneficial to express a fusion protein. The design of constructs therefor and the expression and production of fusion proteins and are well known in the art (Sambrook et al., 1989, *supra*; and Ausubel et al., 1994, *supra*).

Non limiting examples of such fusion proteins include a hemaglutinin fusions and Gluthione-S-transferase (GST) fusions and Maltose binding protein (MBP) fusions. In certain embodiments, it might be beneficial to introduce a protease cleavage site between the two polypeptide sequences which have been fused. Such protease cleavage sites between two heterologously fused polypeptides are well known in the art.

In certain embodiments, it might also be beneficial to fuse the protein of the present invention to signal peptide sequences enabling a secretion of the fusion protein from the host cell. Signal peptides from diverse organisms are well known in the art. Bacterial OmpA and yeast Suc2 are two

non limiting examples of proteins containing signal sequences. In certain embodiments, it might also be beneficial to introduce a linker (commonly known) between the interaction domain and the heterologous polypeptide portion. Such fusion protein find utility in the assays of the present invention as well as for purification purposes, detection purposes and the like.

For certainty, the sequences and polypeptides useful to practice the invention include without being limited thereto mutants, homologs, subtypes, alleles and the like. It shall be understood that generally, the sequences of the present invention should encode a functional (albeit defective) AR. It will be clear to the person of ordinary skill that whether the AR sequence of the present invention, variant, derivative, or fragment thereof retains its function, can be determined by using the teachings and assays of the present invention and the general teachings of the art.

It should be understood that the AR protein of the present invention can be modified, for example by *in vitro* mutagenesis, to dissect the structure-function relationship thereof and permit a better design and identification of modulating compounds. However, some derivative or analogs having lost their biological function may still find utility, for example for raising antibodies. These antibodies could be used for detection or purification purposes. In addition, these antibodies could also act as competitive or non-competitive inhibitor and be found to be modulators of the activity of the AR protein of the present invention.

A host cell or indicator cell has been "transfected" by exogenous or heterologous DNA (e.g. a DNA construct) when such DNA has been introduced inside the cell. The transfecting DNA may or may not be integrated (covalently linked) into chromosomal DNA making up the genome of the cell. In prokaryotes, yeast, and mammalian cells for example, the transfecting DNA may be maintained on a episomal element such as a plasmid. With respect to eukaryotic cells, a stably transfected cell is one in which the transfecting DNA has become integrated into a chromosome so that it is

inherited by daughter cells through chromosome replication. This stability is demonstrated by the ability of the eukaryotic cell to establish cell lines or clones comprised of a population of daughter cells containing the transfecting DNA. Transfection methods are well known in the art (Sambrook et al., 1989, *supra*; Ausubel et al., 1994 *supra*). The use of a mammalian cell as indicator can provide the advantage of furnishing an intermediate factor, which permits for example the interaction of two polypeptides which are tested, that might not be present in lower eukaryotes or prokaryotes. It will be understood that extracts from mammalian cells for example could be used in certain embodiments, to compensate for the lack of certain factors.

In general, techniques for preparing antibodies (including monoclonal antibodies and hybridomas) and for detecting antigens using antibodies are well known in the art (Campbell, 1984, In "Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology", Elsevier Science Publisher, Amsterdam, The Netherlands) and in Harlow et al., 1988 (in: Antibody-A Laboratory Manual, CSH Laboratories). The present invention also provides polyclonal, monoclonal antibodies, or humanized versions thereof, chimeric antibodies and the like which inhibit or neutralize their respective interaction domains and/or are specific thereto.

From the specification and appended claims, the term therapeutic agent should be taken in a broad sense so as to also include a combination of at least two such therapeutic agents. Further, the DNA segments or proteins according to the present invention could be introduced into individuals in a number of ways. For example, cells can be isolated from the afflicted individual, transformed with a DNA construct according to the invention and reintroduced to the afflicted individual in a number of ways. Alternatively, the DNA construct can be administered directly to the afflicted individual. The DNA construct can also be delivered through a vehicle such as a liposome, which can be designed to be targeted to a specific cell type, and engineered to be administered through different routes. For example, an androgen receptor

gene having the genotype associated with low risk of breast cancer could be introduced in cells or in an individual displaying the AR polymorphism associated with high risk of breast cancer.

For administration to humans, the prescribing medical professional will ultimately determine the appropriate form and dosage for a given patient, and this can be expected to vary according to the chosen therapeutic regimen (i.e. DNA construct, protein, cells), the response and condition of the patient as well as the severity of the disease.

Composition within the scope of the present invention should contain the active agent (i.e. molecule, hormone) in an amount effective to achieve the desired therapeutic effect while avoiding adverse side effects. Typically, the nucleic acids in accordance with the present invention can be administered to mammals (i.e. humans) in doses ranging from 0.005 to 1 mg per kg of body weight per day of the mammal which is treated. Pharmaceutically acceptable preparations and salts of the active agent are within the scope of the present invention and are well known in the art (Remington's Pharmaceutical Science, 16th Ed., Mack Ed.). For the administration of polypeptides, antagonists, agonists and the like, the amount administered should be chosen so as to avoid adverse side effects. The dosage will be adapted by the clinician in accordance with conventional factors such as the extent of the disease and different parameters from the patient. Typically, 0.001 to 50 mg/kg/day will be administered to the mammal.

The present invention relates to a kit for assessing a predisposition to breast cancer comprising a determination of the genotype at the AR locus (or a locus in linkage desiquilibrium therewith) using a nucleic acid fragment, a protein or a ligand, or a restriction enzyme in accordance with the present invention. For example, a compartmentalized kit in accordance with the present invention includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allow the efficient transfer of

reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include in one particular embodiment a container which will accept the test sample (DNA protein or cells), a container which contains the primers used in the assay, containers which contain enzymes, containers which contain wash reagents, and containers which contain the reagents used to detect the extension products.

It will be readily recognized by the person of ordinary skill, that the nucleic acid sequences, probes, primers, antibodies and the like of the present invention enabling a detection of the CAG repeat polymorphism of the AR gene of the present invention can be incorporated into anyone of numerous established kit formats which are well known in the art.

Other objects, advantages and features of the present invention will become more apparent upon reading of the following non-restrictive description of preferred embodiments which is exemplary and should not be interpreted as limiting the scope of the present invention.

DESCRIPTION OF THE PREFERRED EMBODIMENT

In accordance with one embodiment of the invention, there is provided a specific model for use in prediction of breast cancer susceptibility and prognosis. The model comprises an androgen receptor gene polymorphism that allows to identify a subset of patients (i.e.women) that are at significantly increased risk of breast cancer as compared to those bearing other variants of this gene.

In accordance with a preferred embodiment of the present invention, a single gene, the androgen receptor gene, has been identified. The polymorphism of this gene is associated with a significant proportion of breast cancer cases in the general population (up to 60% of all cases). Polymorphism of this gene is for example the CAG repeat located in the first exon.

It was thus discovered in accordance with a preferred embodiment of the present invention that testing for this polymorphism in the androgen receptor (AR) gene allows to distinguish between women at lower risk of breast cancer and those at higher risk of the disease.

The present invention will be more readily understood by referring to the following example which is given to illustrate the invention rather than to limit its scope.

EXAMPLE 1

Polymorphism of the CAG repeat of the androgen receptor as a marker for breast cancer susceptibility

In a case control study comparing 262 consecutive cases of breast cancer in women and 465 control women matched for age, polymorphism at the AR gene, namely, the CAG repeat coding for a polyglutamine tract in the 5' part of the AR gene located on chromosome X, was studied. Because of the large number of alleles identified (15 different alleles), these alleles were grouped arbitrarily in categories by size to simplify the analysis and increase the number of individuals in each category. Table 1 presents the frequency of cases and controls in categories of genotypes with the corresponding odds ratio for breast cancer and the computed 95% confidence intervals. The AR gene alleles were called arbitrarily A to E according to their size in CAG repeats, the shortest alleles being A and the longest being called E. The shortest AR gene alleles (corresponding to the polyglutamine stretch) or combinations of short alleles (AA,AB,BB) are the genotypes that show the smallest breast cancer risk. This shows that women with a certain combination of AR gene polymorphisms on their two X chromosomes have a significantly increased risk of developing breast cancer as compared to the category with the smallest risk. In fact, in this cohort 32% of all cases of breast cancer were attributable to variation in the AR gene. This is three to six times the number of breast cancer cases attributable

to the BRCA1 and BRCA2 genes. Indeed, in the cohort studied the 25% of women with the AR genotypes associated with the smallest risk of breast cancer comprised only 19% of all breast cancer cases while the 75% of women having the AR genotypes associated with the highest risks of breast cancer had 81% of all breast cancer cases. In other words, as compared with the general population, for which the risk of breast cancer is of 1:9 women, women with certain AR genotypes had a risk of 1:12 (much lower; i.e. protecting effect) while the other group had a risk of 1:8 (larger). Thus, this novel genetic marker of breast cancer allows to identify a subgroup of women with a risk of breast cancer close to two times larger than the other subgroup.

Table 1

Distribution of cases and controls among females with various AR genotypes

AR genotype	Cases	Controls	Totals
A* + BB	49	134	183
BC to EE	213	331	544
Totals	262	465	727

Odds Ratio (OR) for breast cancer in BC to EE genotypes vs A* and BB = 1.76 (95% confidence interval CI 1.22 to 2.55)
Chi-square = 9.5 p=0.002
Breast cancer risk attributable to AR gene variation = 32% (57 cases / 213 in the others category)

As will be clear to the skilled artisan, the different alleles AR alleles can be grouped differently according to size, and the invention should therefore not be limited to particular groupings. As will be seen in Table 2, groupings of the alleles in three categories instead of 5, still enable a demonstration of the significant association of the AR CAG-repeat polymorphism with breast cancer.

In Table 2, the 15 different alleles were grouped in three different categories (X, Y, and Z) instead of five, in which the shortest alleles are in the X category, and the longest alleles are in the Z category. The six possible genotypes were thus designated as "XX", "XY", "XZ", "YY", "YZ, and "ZZ" genotypes. It is apparent from Table 2 that (CAG)n genotypes were associated with the disease as the genotypes with mid to large numbers of (CAG) repeats were at significantly higher risk of developing the disease as compared to genotypes with shorter (CAG)n tracts (Table 2). Table 2 shows that women with either the YY, YZ or ZZ genotypes had a 2.2-fold increased risk of breast cancer compared to women with the XX or XY genotype, i.e. that women with these later genotypes had only a 1:20 lifetime risk for the disease as compared to a 1:9 risk for those with the larger genotypes.

Table 2

Association of androgen receptor polymorphism with breast cancer

	(CAG)n genotype		
	XX or XY genotype	XZ genotype	YY, YZ or ZZ genotype
Cases	10 (4%)*	28 (11%)*	212 (85%)*
Controls	37 (8%)*	61 (13%)*	355 (78%)*
Odds Ratio (OR)	1.0	1.7	2.2
95% CI for OR (min-max)		0.7 - 3.9	1.1 - 4.5
Lifetime risk of breast cancer	1:20	1:12	1:9

* value in parenthesis represents percentage of total cases or controls
CI Confidence interval expressed with the highest and lowest values.

No significant interaction was observed between AR genotypes and the body mass index (BMI), smoking habits, menopausal status or family history of breast cancer. However, a striking combined influence of the AR genotype with a positive history of breast benign disease (BBD) on the risk of breast cancer was observed (Table 3). Women with a positive history of BBD and AR genotypes combining the large AR alleles (Y or Z) had a relative risk of 3.5 as compared to women with no such history and AR genotypes comprised of smaller alleles. When compared to carriers of XX, XY AR genotypes only (who have the lowest risk of breast cancer) with no history of benign disease, women with the AR-ZZ genotype had an odds ratio of 7.1 for breast cancer (95% CI 2.3 to 22). Interestingly the AR genotype was not associated with a significant risk of breast cancer in women with no history of breast benign disease. The present invention thus also provides as an additional "marker" to

strengthen the prognosis/diagnosis/treatment methods and reagents according to the present invention, a positive history of BBD.

Table 3

Association of breast cancer risk with AR polymorphism and breast benign disease

		AR genotype	
		XX, XY, XZ	YY, YZ, ZZ
Negative history of benign breast disease	Cases	25 (10%)*	131 (52%)*
	Controls	73 (16%)*	288 (64%)*
	Odds Ratio	1.0	1.33
	95% CI for OR (min - max)	--	0.8 - 2.2
	Lifetime risk of breast cancer	1:16	1:12
Positive history of benign breast disease	Cases	13 (5%)*	81 (32%)*
	Controls	25 (6%)*	67 (15%)*
	Odds Ratio	1.5	3.5
	95% CI for OR (min - max)	0.7 - 3.4	2.0 - 6.2
	Lifetime risk of breast cancer	1:11	1:4

* value in parenthesis represents percentage of total cases or controls
CI Confidence interval expressed with the highest and lowest values.

Up to now, no marker displaying such a large odds ratios had been reported for breast cancer. Furthermore, this genetic marker and polymorphisms in the AR gene play a very significant role in breast cancer susceptibility in women, as evidenced by the very significant association demonstrated herein. The present invention also points to alternative therapies for breast cancer aiming at restoring the efficacy of the AR in women with a reduced function of their AR genes due to the variant genotypes that they carry. The described assays of the present invention could enables the identification of such therapies.

Although the present invention has been described hereinabove by way of preferred embodiments thereof, it can be modified, without departing from the spirit and nature of the subject invention as defined in the appended claims.

SEQUENCE LISTING

SEQ ID NO:1 5'-TCCAGAATCTGTTCCAGAGCGTGC-3

SEQ ID NO:2 5'-GCTGTGAAGGTTGCTGTTCCTCAT-3'

WHAT IS CLAIMED IS:

1. A method of determining an individual's predisposition to breast cancer, development of breast cancer and/or responsiveness to therapy for breast cancer, said method comprising the step of determining the androgen receptor genotype of the individual, thereby determining an individual's predisposition to breast cancer, development of breast cancer and/or responsiveness to therapy for breast cancer.

2. The method of claim 1, wherein the androgen receptor genotype is determined using a nucleic acid variant in linkage disequilibrium with a CAG repeat in an androgen receptor gene.

3. The method of claim 2, wherein the androgen receptor genotype is determined by determining the number of CAG repeat within the androgen receptor gene

4. The method of claim 3, which further comprises a step of amplifying a segment of the androgen receptor using polymerase chain reaction.

5. The method of claim 4, wherein a pair of primers derived from a nucleic acid sequence of the androgen receptor gene or flanking said gene is used in the polymerase chain reaction.

6. The method of claim 5, wherein the segment of the androgen receptor gene is amplified using a pair of primers as follows:

5'-TCCAGAATCT GTTCCAGAGC GTGC-3' SEQ ID NO:1; and
5'-GCTGTGAAGG TTGCTGTTCC TCAT-3' SEQ ID NO:2.

7. An assay for screening and selecting an agent which modulates breast cancer predisposition comprising:

a) a recombinant androgen receptor gene or functional fragment thereof, which comprisesthe CAG repeat thereof, or a marker in linkage disequilibrium therewith; and

b) assaying a function of said androgen receptor;

wherein an allele which modulates said function of an androgen receptor can be selected,

and wherein a modulation of a function of said androgen receptor is associated with a modulation of said breast cancer predisposition, whereby short CAG repeat of said AR positively modulates AR receptor function, thereby leading to breast cancer protection.

8. An assay for screening and selecting an agent which modulates breast cancer predisposition comprising:

a) an expression vector comprising a promoter operably linked to a reporter gene, said promoter comprising an androgen response element, said response element affecting the activity of said promoter upon binding thereto of androgen or analog thereof;

b) a cell expressing a chosen allele of an androgen receptor and harboring said vector of a);

c) submitting said cell to at least one agent; and

d) assaying a level of said reporter gene;

whereby an agent can be selected when the level of said reporter gene is significantly modulated by the presence of said agent.

9. A method of using specific alleles of the androgen receptor gene, or a variant, equivalent, or mutation thereof which shows linkage disequilibrium therewith, to set-up a screening assay for agents destined to modulate androgen receptor function for the purpose of identifying agents involved in breast cancer.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12Q 1/68

A3

(11) International Publication Number: WO 00/15834

(43) International Publication Date: 23 March 2000 (23.03.00)

(21) International Application Number: PCT/CA99/00852

(22) International Filing Date: 15 September 1999 (15.09.99)

(30) Priority Data:
60/100,311 15 September 1998 (15.09.98) US

(71) Applicant *(for all designated States except US)*: SIGNALGENE INC. [CA/CA]; 8475 Christophe Colomb Avenue, Montreal, Quebec H2M 2N9 (CA).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: ROUSSEAU, François [CA/CA]; 2540 Lalonde, Sainte-Foy, Quebec G1W 1M7 (CA).

(74) Agents: DUBUC, Jean, H. et al.; Goudreau Gage Dubuc & Martineau Walker, The Stock Exchange Tower, Suite 3400, 800 Place Victoria, P.O. Box 242, Montreal, Quebec H4Z 1E9 (CA).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(88) Date of publication of the international search report:
25 May 2000 (25.05.00)

(54) Title: MARKER AT THE ANDROGEN RECEPTOR GENE FOR DETERMINING BREAST CANCER SUSCEPTIBILITY

(57) Abstract

The present invention relates to a method of determining an individual's predisposition to breast cancer, development of breast cancer, protection against breast cancer and/or responsiveness to therapy for breast cancer. The method comprises the step of determining the androgen receptor genotype at the CAG repeat locus of an individual, or at a locus in linkage disequilibrium with the CAG repeat locus, thereby determining an individual's predisposition to breast cancer, development of breast cancer, protection against breast cancer and/or responsiveness to therapy for breast cancer.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania
AM	Armenia
AT	Austria
AU	Australia
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CU	Cuba
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GH	Ghana
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IL	Israel
IS	Iceland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakstan
LC	Saint Lucia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LS	Lesotho
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
MK	The former Yugoslav Republic of Macedonia
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TM	Turkmenistan
TR	Turkey
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam
YU	Yugoslavia
ZW	Zimbabwe

INTERNATIONAL SEARCH REPORT

Inter[redacted] Application No
PCT/CA 99/00852

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12Q C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	ELHAJI Y ET AL: "The polymorphic CAG repeat of the androgen receptor and female breast cancer" AMERICAN JOURNAL OF HUMAN GENETICS, vol. 61, no. 4 suppl, 28 October 1997 (1997-10-28) - 1 November 1997 (1997-11-01), page A64 XP000884969 ⇒ see abstract 346	1-5

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
8 March 2000	16/03/2000

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340–2040, Tx. 31 651 epo nl, Fax: (+31–70) 340–3016	Osborne, H

1

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	REBBECK TR ET AL: "Modification of breast cancer risk in BRCA1 mutation carriers by the androgen receptor CAG repeat polymorphism" PROCEEDINGS OF THE AMERICAN ASSOCIATION FOR CANCER RESEARCH ANNUAL MEETING, vol. 39, 28 March 1998 (1998-03-28) - 1 April 1998 (1998-04-01), page 366 XP000884978 see abstract 2493	1-5
A	WO 97 17469 A (DANA FARBER CANCER INST INC ;BRIGHAM & WOMENS HOSPITAL (US)) 15 May 1997 (1997-05-15)	1-6
A	HALL RE ET AL: "Regulation of androgen receptor gene expression by steroids and retinoic acid in human breast cancer cells" INTERNATIONAL JOURNAL OF CANCER, vol. 52, no. 5, November 1992 (1992-11), pages 778-84, XP000884982 see abstract	7,8
A	WO 98 05797 A (LAMPARSKI HENRY G ;CALYDON (US); SCHUUR ERIC R (US); YU DE CHAO (U) 12 February 1998 (1998-02-12) page 8, line 4 - line 13	7-9

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

Inter[redacted] Application No
PCT/CA 99/00852

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9717469 A	15-05-1997	AU 7722696 A	29-05-1997
WO 9805797 A	12-02-1998	US 5783435 A	21-07-1998
		AU 3972997 A	25-02-1998
		CA 2262438 A	12-02-1998
		EP 0918884 A	02-06-1999

INTERNATIONAL SEARCH REPORT

International Application No
P[redacted]CA 99/00852

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12Q C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	ELHAJI Y ET AL: "The polymorphic CAG repeat of the androgen receptor and female breast cancer" AMERICAN JOURNAL OF HUMAN GENETICS, vol. 61, no. 4 suppl, 28 October 1997 (1997-10-28) - 1 November 1997 (1997-11-01), page A64 XP000884969 see abstract 346 —/—	1-5

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
8 March 2000	16/03/2000

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340–2040, Tx. 31 651 epo nl, Fax: (+31–70) 340–3016	Osborne, H

1

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	REBBECK TR ET AL: "Modification of breast cancer risk in BRCA1 mutation carriers by the androgen receptor CAG repeat polymorphism" PROCEEDINGS OF THE AMERICAN ASSOCIATION FOR CANCER RESEARCH ANNUAL MEETING, vol. 39, 28 March 1998 (1998-03-28) - 1 April 1998 (1998-04-01), page 366 XP000884978 see abstract 2493	1-5
A	WO 97 17469 A (DANA FARBER CANCER INST INC ;BRIGHAM & WOMENS HOSPITAL (US)) 15 May 1997 (1997-05-15)	1-6
A	HALL RE ET AL: "Regulation of androgen receptor gene expression by steroids and retinoic acid in human breast cancer cells" INTERNATIONAL JOURNAL OF CANCER, vol. 52, no. 5, November 1992 (1992-11), pages 778-84, XP000884982 see abstract	7,8
A	WO 98 05797 A (LAMPARSKI HENRY G ;CALYDON (US); SCHUUR ERIC R (US); YU DE CHAO (U) 12 February 1998 (1998-02-12) page 8, line 4 - line 13	7-9

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/CA 99/00852

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9717469 A	15-05-1997	AU 7722696 A	29-05-1997
WO 9805797 A	12-02-1998	US 5783435 A	21-07-1998
		AU 3972997 A	25-02-1998
		CA 2262438 A	12-02-1998
		EP 0918884 A	02-06-1999

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
7 February 2002 (07.02.2002)

PCT

(10) International Publication Number
WO 02/10452 A2

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number: PCT/US01/23834

(22) International Filing Date: 27 July 2001 (27.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/221,074 27 July 2000 (27.07.2000) US

(71) Applicant *(for all designated States except US)*: **UNIVERSITY OF ROCHESTER** [US/US]; 601 Elmwood Avenue, Box 706, Rochester, NY 14642 (US).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **CHANG, Chawnshang** [US/US]; 19 Sandy Lane, Pittsford, NY 14534 (US).

(74) Agents: **SPRATT, Gwendolyn, D.** et al.; Needle & Rosenberg, P.C., Suite 1200, The Candler Building, 127 Peachtree Street, N.E., Atlanta, GA 30303-1811 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/10452 A2

(54) Title: METHODS AND COMPOSITIONS FOR PREDICTING PROSTATE CANCER

(57) Abstract: Disclosed are compositions and methods for assessing risk in a subject for prostate cancer.

METHODS AND COMPOSITIONS FOR PREDICTING PROSTATE CANCER

This application claims priority to United States Provisional Application No 60/221,074 filed on July 27, 2000, which application is herein incorporated by reference in its entirety.

I. BACKGROUND OF THE INVENTION

The incidence of clinical prostate cancer differs substantially between ethnic groups, with African Americans having a 10- to 40-fold higher incidence than Asians (1-3AR). Such disparity in incidence of clinical prostate cancer cannot be explained entirely by population differences in screening. An earlier study shows that after adjustment for screening, there is still a 3- to 4-fold difference in incidence rates between U.S. and Japanese men, whose rates are among the highest in Asians (4AR). Despite the dramatic racial variation in clinical prostate cancer incidence, the prevalence of latent carcinoma appears to be similar across populations (5AR), suggesting that there exists differences in factors (either genetic or environmental) that promote the progression of microscopic tumors to clinically overt carcinoma.

The growth, differentiation, and proliferation of prostatic cells are regulated by androgens (6AR). The biological effects of androgens are mediated through binding to the intracellular androgen receptor (AR), which in turn regulates the transcription of target genes with the assistance of transcriptional coactivators (7AR).

II. SUMMARY OF THE INVENTION

In accordance with the purposes of this invention, as embodied and broadly described herein, this invention, in one aspect, relates compositions and methods for assessing prostate cancer risk.

Additional advantages of the invention will be set forth in part in the description which follows, and in part will be obvious from the description, or may be learned by practice of the invention. The advantages of the invention will be realized and attained by means of the elements and combinations particularly pointed out in the appended claims. It is to be understood that both the foregoing general description and the following detailed description are exemplary and

explanatory only and are not restrictive of the invention, as claimed.

III. BRIEF DESCRIPTION OF THE DRAWINGS

The accompanying drawings, which are incorporated in and constitute a part of this specification, illustrate several embodiments of the invention and together with the description, serve to explain the principles of the invention.

Figure 1. (AR) shows the percent distribution of the number of CAG repeats.

IV. DETAILED DESCRIPTION

The present invention may be understood more readily by reference to the following detailed description of preferred embodiments of the invention and the Examples included therein and to the Figures and their previous and following description.

Before the present compounds, compositions, articles, devices, and/or methods are disclosed and described, it is to be understood that this invention is not limited to specific synthetic methods, specific recombinant biotechnology methods unless otherwise specified, or to particular reagents unless otherwise specified, as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting.

A. Definitions

As used in the specification and the appended claims, the singular forms "a," "an" and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a pharmaceutical carrier" includes mixtures of two or more such carriers, and the like.

Ranges may be expressed herein as from "about" one particular value, and/or to "about" another particular value. When such a range is expressed, another embodiment includes from the one particular value and/or to the other particular value. Similarly, when values are expressed as approximations, by use of the antecedent "about," it will be understood that the particular value forms another embodiment. It will be further understood that the endpoints of each of the ranges

are significant both in relation to the other endpoint, and independently of the other endpoint.

In this specification and in the claims which follow, reference will be made to a number of terms which shall be defined to have the following meanings:

"Optional" or "optionally" means that the subsequently described event or circumstance may or may not occur, and that the description includes instances where said event or circumstance occurs and instances where it does not.

"Primers" are a subset of probes which are capable of supporting some type of enzymatic manipulation and which can hybridize with a target nucleic acid such that the enzymatic manipulation can occur. A primer can be made from any combination of nucleotides or nucleotide derivatives or analogs available in the art which do not interfere with the enzymatic manipulation.

"Probes" are molecules capable of interacting with a target nucleic acid, typically in a sequence specific manner, for example through hybridization. The hybridization of nucleic acids is well understood in the art and discussed herein. Typically a probe can be made from any combination of nucleotides or nucleotide derivatives or analogs available in the art.

B. Compositions

Disclosed are the components to be used to prepare the disclosed compositions as well as the compositions themselves and to be used within the methods disclosed herein. These and other materials are disclosed herein, and it is understood that when combinations, subsets, interactions, groups, etc. of these materials are disclosed that while specific reference of each various individual and collective permutation of these compounds may not be explicitly disclosed, each is specifically contemplated and described herein. For example, if a particular AIB1 protein or gene is disclosed and discussed and a number of modifications that can be made to a number of molecules including the AIB1 protein or gene are discussed, specifically contemplated is each and every combination and permutation of AIB1 protein or gene and the modifications that are possible unless specifically indicated

to the contrary. Thus, if a class of molecules A, B, and C are disclosed as well as a class of molecules D, E, and F and an example of a combination molecule, A-D is disclosed, then even if it each is not individually recited each is individually and collectively contemplated meaning combinations, A-E, A-F, B-D, B-E, B-F, C-D, C-E, and C-F are considered disclosed. Likewise, any subset or combination of these is also disclosed. Thus, for example, the sub-group of A-E, B-F, and C-E would be considered disclosed. This concept applies to all aspects of this application including, but not limited to, steps in methods of making and using the disclosed compositions. Thus, if there are a variety of additional steps that can be performed it is understood that each of these additional steps can be performed with any specific embodiment or combination of embodiments of the disclosed methods.

It is understood that one way to define any known variants and derivatives or those that might arise, of the disclosed genes and proteins herein is through defining the variants and derivatives in terms of homology to specific known sequences. For example, SEQ ID NO:18 sets forth a particular sequence of an AIB1 gene and SEQ ID NO:19 sets forth a particular sequence of the protein encoded by SEQ ID NO:18, an AIB1 protein. Specifically disclosed are variants of these and other genes and proteins herein disclosed which have at least, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 percent homology to the stated sequence. Those of skill in the art readily understand how to determine the homology of two proteins or nucleic acids, such as genes. For example, the homology can be calculated after aligning the two sequences so that the homology is at its highest level.

Another way of calculating homology can be performed by published algorithms. Optimal alignment of sequences for comparison may be conducted by the local homology algorithm of Smith and Waterman Adv. Appl. Math. 2: 482 (1981), by the homology alignment algorithm of Needleman and Wunsch, J. MoL Biol. 48: 443 (1970), by the search for similarity method of Pearson and Lipman, Proc. Natl. Acad. Sci. U.S.A. 85: 2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics

Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by inspection.

The same types of homology can be obtained for nucleic acids by for example the algorithms disclosed in Zuker, M. *Science* 244:48-52, 1989, Jaeger et al. *Proc. Natl. Acad. Sci. USA* 86:7706-7710, 1989, Jaeger et al. *Methods Enzymol.* 183:281-306, 1989 which are herein incorporated by reference for at least material related to nucleic acid alignment.

The term hybridization typically means a sequence driven interaction between at least two nucleic acid molecules, such as a primer or a probe and a gene. Sequence driven interaction means an interaction that occurs between two nucleotides or nucleotide analogs or nucleotide derivatives in a nucleotide specific manner. For example, G interacting with C or A interacting with T are sequence driven interactions. Typically sequence driven interactions occur on the Watson-Crick face or Hoogsteen face of the nucleotide. The hybridization of two nucleic acids is affected by a number of conditions and parameters known to those of skill in the art. For example, the salt concentrations, pH, and temperature of the reaction all affect whether two nucleic acid molecules will hybridize.

Parameters for selective hybridization between two nucleic acid molecules are well known to those of skill in the art. For example, in some embodiments selective hybridization conditions can be defined as stringent hybridization conditions. For example, stringency of hybridization is controlled by both temperature and salt concentration of either or both of the hybridization and washing steps. For example, the conditions of hybridization to achieve selective hybridization may involve hybridization in high ionic strength solution (6X SSC or 6X SSPE) at a temperature that is about 12-25°C below the Tm (the melting temperature at which half of the molecules dissociate from their hybridization partners) followed by washing at a combination of temperature and salt concentration chosen so that the washing temperature is about 5°C to 20°C below the Tm. The temperature and salt conditions are readily determined empirically in preliminary experiments in which samples of reference DNA immobilized on filters

are hybridized to a labeled nucleic acid of interest and then washed under conditions of different stringencies. Hybridization temperatures are typically higher for DNA-RNA and RNA-RNA hybridizations. The conditions can be used as described above to achieve stringency, or as is known in the art. (Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989; Kunkel et al. Methods Enzymol. 1987:154:367, 1987 which is herein incorporated by reference for material at least related to hybridization of nucleic acids). A preferable stringent hybridization condition for a DNA:DNA hybridization can be at about 68°C (in aqueous solution) in 6X SSC or 6X SSPE followed by washing at 68°C. Stringency of hybridization and washing, if desired, can be reduced accordingly as the degree of complementarity desired is decreased, and further, depending upon the G-C or A-T richness of any area wherein variability is searched for. Likewise, stringency of hybridization and washing, if desired, can be increased accordingly as homology desired is increased, and further, depending upon the G-C or A-T richness of any area wherein high homology is desired, all as known in the art.

Another way to define selective hybridization is by looking at the amount (percentage) of one of the nucleic acids bound to the other nucleic acid. For example, in some embodiments selective hybridization conditions would be when at least, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 percent of the limiting nucleic acid is bound to the non-limiting nucleic acid. Typically, the non-limiting primer is in for example, 10 or 100 or 1000 fold excess. This type of assay can be performed at under conditions where both the limiting and non-limiting primer are for example, 10 fold or 100 fold or 1000 fold below their k_d, or where only one of the nucleic acid molecules is 10 fold or 100 fold or 1000 fold or where one or both nucleic acid molecules are above their k_d.

Another way to define selective hybridization is by looking at the percentage of primer that gets enzymatically manipulated under conditions where hybridization is required to promote the desired enzymatic manipulation. For example, in some

embodiments selective hybridization conditions would be when at least, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 percent of the primer is enzymatically manipulated under conditions which promote the enzymatic manipulation, for example if the enzymatic manipulation is DNA extension, then selective hybridization conditions would be when at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 percent of the primer molecules are extended. Preferred conditions also include those suggested by the manufacturer or indicated in the art as being appropriate for the enzyme performing the manipulation.

1. AIB1/SRC3

The AIB1 protein (Amplified In Breast cancer 1, also known as Steroid Receptor Coactivator-3, or SRC-3), encoded by the *AIB1/SRC-3* gene located on chromosome 20 (20q12), is an AR coactivator and a member of the steroid receptor coactivator family, which interacts with members of the nuclear hormone receptor family (14,15 AIB1). Like the AR protein, the AIB1/SRC-3 coactivator contains a stretch of glutamine residues encoded by a variable-length track of CAG/CAA repeats in the *AIB1/SRC-3* gene.

a) Sequences

There are a variety of sequences related to the AIB1 gene having the following Genbank Accession Numbers: NT_011362, XM_030033, XM_030032, XM_009483, XM_030031, XM_030034, AL353777, AL034418, AL021394, AY008258, AF322224, NM_008679, NM_006534, AF012108, and AF044080, these sequences and others are herein incorporated by reference in their entireties as well as for individual subsequences contained therein.

One particular sequence set forth in SEQ ID NO:18 and having Genbank accession number XM_030032 is used herein, as an example, to exemplify the disclosed compositions and methods. It is understood that the description related to this sequence is applicable to any sequence related to AIB1 unless specifically

indicated otherwise. Those of skill in the art understand how to resolve sequence discrepancies and differences and to adjust the compositions and methods relating to a particular sequence to other related sequences (i.e. sequences of AIB1). Primers and/or probes can be designed for any AIB1 sequence given the information disclosed herein and known in the art.

b) CAG/CAA region

The CAG/CAA region of the AIB1 gene is located in the coding region of the AIB1 gene. For example, the CAG/CAA region in SEQ ID NO:18 can be defined by the region from nucleotide 3930 to nucleotide 4016. In certain embodiments of the disclosed compositions and methods, this represents the CAG/CAA region of the AIB1 gene. However, it is understood that various mutations, alterations, or other genetic variation including allelic variation can occur in certain individuals and those of skill in the art understand how to locate this region within any given AIB1 gene variant. Thus, for example, if a nucleic acid is amplified from SEQ ID NO:18 or any variant of the AIB1 gene, that contains only only a fragment of the AIB1 gene, for example, a fragment of 1000 nucleotides, it is understood that the CAG/CAA region could be located within this molecule if it is included in the molecule, in whole or in part.

c) Primers and probes

Disclosed are compositions including primers and probes, which are capable of interacting with the AIB1 gene as disclosed herein. In certain embodiments the primers are used to support DNA amplification reactions. Typically the primers will be capable of being extended in a sequence specific manner. Extension of a primer in a sequence specific manner includes any methods wherein the sequence and/or composition of the nucleic acid molecule to which the primer is hybridized or otherwise associated directs or influences the composition or sequence of the product produced by the extension of the primer. Extension of the primer in a sequence specific manner therefore includes, but is not limited to, PCR, DNA sequencing, DNA extension, DNA polymerization, RNA transcription, or reverse transcription. Techniques and conditions that amplify the primer in a sequence

specific manner are preferred. In certain embodiments the primers are used for the DNA amplification reactions, such as PCR or direct sequencing. It is understood that in certain embodiments the primers can also be extended using non-enzymatic techniques, where for example, the nucleotides or oligonucleotides used to extend the primer are modified such that they will chemically react to extend the primer in a sequence specific manner. Typically the disclosed primers hybridize with the AIB1 gene or region of the AIB1 gene or they hybridize with the complement of the AIB1 gene or complement of a region of the AIB1 gene.

Disclosed are primers that are capable of amplifying the CAG/CAA region of the AIB1 gene. In certain embodiments the primers amplify the CAG/CAA region of the AIB1 gene from nucleotide 3930 to nucleotide 4016 of the sequence set forth in SEQ ID NO:18. In certain embodiments the primers are "outside" the CAG/CAA region of the AIB1 gene. By outside the region it is intended to indicate that no region of the primer is intended to interact directly with the CAG/CAA region. For example, a primer outside the CAG/CAA region of the sequence set forth in SEQ ID NO:18 could hybridize with nucleotide 4017 to nucleotide 4037 of SEQ ID NO:18, but it would not be intended to hybridize with nucleotide 4016, and under the conditions designed for the enzymatic manipulation of the primer, it would not appreciably interact with nucleotide 4016. In other embodiments the primers are designed to interact with one or more nucleotides considered to be part of the CAG/CAA region, which primers herein are referred to as "inside" primers. Thus, an inside primer can include a primer region that can under conditions appropriate for the desired enzymatic manipulation hybridize with one or more nucleotides considered within the CAG/CAA region and contain a region that hybridizes with nucleotides that are not considered part of the CAG/CAA region. Thus, in one embodiment an inside primer would interact with the nucleotide at position 4016 of SEQ ID NO:18.

The size of the primers for interaction with the AIB1 gene in certain embodiments can be any size that supports the desired enzymatic manipulation of the primer, such as DNA amplification. A typical AIB1 primer would be at least 6,

7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1250, 1500, 1750, 2000, 2250, 2500, 2750, 3000, 3500, or 4000 nucleotides long.

In other embodiments an AIB1 primer can be less than or equal to 6, 7, 8, 9, 10, 11, 12 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1250, 1500, 1750, 2000, 2250, 2500, 2750, 3000, 3500, or 4000 nucleotides long.

In certain embodiments the primers are designed such that they are outside primers whose nearest point of interaction with the AIB1 gene is within 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 125, 150, 175, or 200 nucleotides of the outermost defining nucleotide of the CAG/CAA region or complement of the CAG/CAA region.

In certain embodiments the primers are designed such that they are outside primers whose nearest point of interaction with the AIB1 gene is at least 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 125, 150, 175, or 200 nucleotides away from the outermost

defining nucleotide of the CAG/CAA region or complement of the CAG/CAA region.

For example, with respect to the AIB1 gene set forth in SEQ ID NO:18, certain embodiments of the primers would be designed such that they are outside primers whose nearest point of interaction with the AIB1 gene occurs at position 4017, 4018, 4019, 4020, 4021, 4022, 4023, 4024, 4025, 4026, 4027, 4028, 4029, 4030, 4031, 4032, 4033, 4034, 4035, 4036, 4037, 4038, 4039, 4040, 4041, 4042, 4043, 4044, 4045, 4046, 4047, 4048, 4049, 4050, 4051, 4052, 4053, 4054, 4055, 4056, 4057, 4058, 4059, 4060, 4061, 4062, 4063, 4064, 4065, 4066, 4067, 4068, 4069, 4070, 4071, 4072, 4073, 4074, 4075, 4076, 4077, 4078, 4079, 4080, 4081, 4082, 4083, 4084, 4085, 4086, 4087, 4088, 4089, 4090, 4091, 4092, 4093, 4094, 4095, 4096, 4097, 4098, 4099, 4100, 4101, 4102, 4103, 4104, 4105, 4106, 4107, 4108, 4109, 4110, 4111, 4112, 4113, 4114, 4115, 4116, 4117, 4142, 4167, 4192, 4217 of SEQ ID NO:18.

The primers for the AIB1 gene typically will be used to produce an amplified DNA product that contains the CAG/CAA region of the AIB1 gene. In general, typically the size of the product will be such that the size can be accurately determined to within 3, or 2 or 1 nucleotides.

In certain embodiments this product is at least 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1250, 1500, 1750, 2000, 2250, 2500, 2750, 3000, 3500, or 4000 nucleotides long.

In other embodiments the product is less than or equal to 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90,

91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1250, 1500, 1750, 2000, 2250, 2500, 2750, 3000, 3500, or 4000 nucleotides long.

2. Androgen receptor

Androgen receptor (AR) plays a key role in intraprostatic androgenic action. Within the prostate gland, testosterone is converted into dihydrotestosterone (DHT), a more potent androgen. DHT then binds to the AR to form an intracellular DHT-AR complex, which in turn modulates prostatic target genes to induce proliferation.

The AR protein, consisting of 918 amino acids and encoded singly by the AR gene located on the X chromosome (Xq11-12), has three major functional domains: a transactivating amino-terminal domain, a DNA binding domain, and a ligand (steroid) binding domain (8AR). The open reading frame of the AR gene is separated over eight exons and has a length of 2,730 base pair (bp). The sequence encoding the large amino-terminal transactivating domain is found in the first exon; the DNA binding domain is encoded by exons 2 and 3; and the information for the ligand binding domain is distributed over exons 4 to 8 (8AR).

The first exon of the AR gene contains two polymorphic trinucleotide repeat segments that encode polyglutamine and polyglycine tracts localized in the N-terminal transactivation domain of the AR protein. The polyglutamine tract is encoded by a CAG trinucleotide repeat, and the polyglycine stretch by a GGN repeat. The number of CAG repeats ranges from about 8 to 35 repeats in normal individuals. Longer CAG repeat lengths appear to result in reduced AR transcriptional activity both *in vivo* and *in vitro* (9,10AR). Otherwise healthy men whose androgen receptor has a CAG repeat length at the long end of the normal range (>28) have an increased incidence of impaired spermatogenesis and infertility (11AR), conditions that are extremely androgen-dependent (12AR). Expansion of the CAG repeat length to over 40 repeats is related to a rare neuromuscular disorder, spinal and bulbar muscular atrophy (Kennedy syndrome), which is also associated with androgen insensitivity, decreased virilization, testicular atrophy, reduced sperm

production, and infertility (13-15AR). Together, these clinical data suggest that a longer CAG repeat length decreases the functional competence of AR.

The length of the polyglycine (GGN) tract varies from about 10 to 30 repeats. The functional consequences of variation in the GGN tract are less clear. Deletion of the polyglycine tract reduces AR transcriptional activity by about 30% in transient transfection assays (16AR), although there is no significant correlation between polyglycine tract length and infertility (11AR).

a) Sequences

There are a variety of sequences related to the androgen receptor gene having the following Genbank Accession Numbers: NM_000044 or L49399 are two examples. These sequences and others available on Genbank are herein incorporated by reference in their entireties as well as for individual subsequences contained therein.

One particular sequence set forth in SEQ ID NO:20 and having Genbank accession number NM_000044 is used herein, as an example, to exemplify the disclosed compositions and methods. It is understood that the description related to this sequence is applicable to any sequence related to androgen receptor unless specifically indicated otherwise. Those of skill in the art understand how to resolve sequence discrepancies and differences and to adjust the compositions and methods relating to a particular sequence to other related sequences. Primers and/or probes can be designed for any androgen receptor sequence given the information disclosed herein and known in the art.

b) CAG/CAA region

The CAG/CAA region in SEQ ID NO:20 can be defined by the region from nucleotide 1286 to nucleotide 1348. In certain embodiments of the disclosed compositions and methods, this represents the CAG/CAA region of the androgen receptor gene. However, it is understood that various mutations, alterations, or other genetic variation including allelic variation can occur in certain individuals and those of skill in the art understand how to locate this region within any given

androgen receptor gene variant. Thus, for example, if a nucleic acid is amplified from SEQ ID NO:20 or any variant of the androgen receptor gene, that contains only 1000 nucleotides, it is understood that the CAG/CAA region could be located within this molecule if it is included in the molecule, in whole or in part.

c) GGN region

The GGN region in SEQ ID NO:20 can be defined by the region from nucleotide 2459 to nucleotide 2530. In certain embodiments of the disclosed compositions and methods, this represents the GGN region of the androgen receptor gene. However, it is understood that various mutations, alterations, or other genetic variation including allelic variation can occur in certain individuals and those of skill in the art understand how to locate this region within any given androgen receptor gene variant. Thus, for example, if a nucleic acid is amplified from SEQ ID NO:20 or any variant of the androgen receptor gene, that contains only 1000 nucleotides, it is understood that the GGN region could be located within this molecule if it is included in the molecule, in whole or in part.

d) Primers and probes

Disclosed are compositions including primers and probes, which are capable of interacting with the androgen receptor gene as disclosed herein. In certain embodiments the primers are used to support DNA amplification reactions. Typically the primers will be capable of being extended in a sequence specific manner. Extension of a primer in a sequence specific manner includes any methods wherein the sequence and/or composition of the nucleic acid molecule to which the primer is hybridized or otherwise associated directs or influences the composition or sequence of the product produced by the extension of the primer. Extension of the primer in a sequence specific manner therefore includes, but is not limited to, PCR, DNA sequencing, DNA extension, DNA polymerization, RNA transcription, or reverse transcription. Techniques and conditions that amplify the primer in a sequence specific manner are preferred. In certain embodiments the primers are used for the DNA amplification reactions, such as PCR or direct sequencing. It is understood that in certain embodiments the primers can also be extended using non-

enzymatic techniques, where for example, the nucleotides or oligonucleotides used to extend the primer are modified such that they will chemically react to extend the primer in a sequence specific manner. Typically the disclosed primers hybridize with the androgen receptor gene or region of the androgen receptor gene or they hybridize with the complement of the androgen receptor gene or complement of a region of the androgen receptor gene.

Disclosed are primers that are capable of amplifying the CAG/CAA and/or GGN region of the androgen receptor gene. In certain embodiments the primers amplify the CAG/CAA region of the androgen receptor gene from nucleotide 1286 to nucleotide 1348 of the sequence set forth in SEQ ID NO:20. In certain embodiments the primers are "outside" the CAG/CAA region of the androgen receptor gene. By outside the region it is intended to indicate that no region of the primer is intended to interact directly with the CAG/CAA region. For example, a primer outside the CAG/CAA region of the sequence set forth in SEQ ID NO:20 could hybridize with nucleotide 1349 to nucleotide 1369 of SEQ ID NO:20, but it would not be intended to hybridize with nucleotide 1348, and under the conditions designed for the enzymatic manipulation of the primer, it would not appreciably interact with nucleotide 1348. In other embodiments the primers are designed to interact with one or more nucleotides considered to be part of the CAG/CAA region, which primers herein are referred to as "inside" primers. Thus, an inside primer can include a primer region that can under conditions appropriate for the desired enzymatic manipulation hybridize with one or more nucleotides considered within the CAG/CAA region and contain a region that hybridizes with nucleotides that are not considered part of the CAG/CAA region. Thus, in one embodiment an inside primer would interact with the nucleotide at position 1348 of SEQ ID NO:20.

In certain embodiments the primers amplify the GGN region of the androgen receptor gene from nucleotide 2459 to nucleotide 2530 of the sequence set forth in SEQ ID NO:20. In certain embodiments the primers are "outside" the GGN region of the androgen receptor gene. By outside the region it is intended to indicate that no region of the primer is intended to interact directly with the GGN region. For

example, a primer outside the GGN region of the sequence set forth in SEQ ID NO:20 could hybridize with nucleotide 2531 to nucleotide 2551 of SEQ ID NO:20, but it would not be intended to hybridize with nucleotide 2530, and under the conditions designed for the enzymatic manipulation of the primer, it would not appreciably interact with nucleotide 2530. In other embodiments the primers are designed to interact with one or more nucleotides considered to be part of the GGN region, which primers herein are referred to as "inside" primers. Thus, an inside primer can include a primer region that can under conditions appropriate for the desired enzymatic manipulation hybridize with one or more nucleotides considered within the GGN region and contain a region that hybridizes with nucleotides that are not considered part of the GGN region. Thus, in one embodiment an inside primer would interact with the nucleotide at position 2530 of SEQ ID NO:20.

The size of the primers for interaction with the androgen receptor gene in certain embodiments can be any size that supports the desired enzymatic manipulation of the primer, such as DNA amplification. A typical androgen receptor primer would be at least 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1250, 1500, 1750, 2000, 2250, 2500, 2750, 3000, 3500, or 4000 nucleotides long.

In other embodiments an androgen receptor primer can be less than or equal to 6, 7, 8, 9, 10, 11, 12 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1250, 1500,

1750, 2000, 2250, 2500, 2750, 3000, 3500, or 4000 nucleotides long.

In certain embodiments the primers are designed such that they are outside primers whose nearest point of interaction with the androgen receptor gene is within 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 125, 150, 175, or 200 nucleotides of the outermost defining nucleotide of the CAG/CAA and/or GGN region or complement of the CAG/CAA and/or GGN region.

In certain embodiments the primers are designed such that they are outside primers whose nearest point of interaction with the androgen receptor gene is at least 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 125, 150, 175, or 200 nucleotides away from the outermost defining nucleotide of the CAG/CAA and/or GGN region or complement of the CAG/CAA and/or GGN region.

For example, with respect to the androgen receptor gene set forth in SEQ ID NO:20, certain embodiments of the primers would be designed such that they are outside primers whose nearest point of interaction with the androgen receptor gene occurs at position 1349, 1350, 1351, 1352, 1353, 1354, 1355, 1356, 1357, 1358, 1359, 1360, 1361, 1362, 1363, 1364, 1365, 1366, 1367, 1368, 1369, 1370, 1371, 1372, 1373, 1374, 1375, 1376, 1377, 1378, 1379, 1380, 1381, 1382, 1383, 1384, 1385, 1386, 1387, 1388, 1389, 1390, 1391, 1392, 1393, 1394, 1395, 1396, 1397, 1398, 1399, 1400, 1401, 1402, 1403, 1404, 1405, 1406, 1407, 1408, 1409, 1410, 1411, 1412, 1413, 1414, 1415, 1416, 1417, 1418, 1419, 1420, 1421, 1422, 1423, 1424, 1425, 1426, 1427, 1428, 1429, 1430, 1431, 1432, 1433, 1434, 1435, 1436, 1437, 1438, 1439, 1440, 1441, 1442, 1443, 1444, 1445, 1446, 1447, 1448, 1449,

1474, 1499, 1524, or 1549 of SEQ ID NO:20.

The primers for the androgen receptor gene typically will be used to produce an amplified DNA product that contains the CAG/CAA and/or GGN region of the androgen receptor gene. In general, typically the size of the product will be such that the size can be accurately determined to within 3, or 2 or 1 nucleotides.

In certain embodiments this product is at least 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1250, 1500, 1750, 2000, 2250, 2500, 2750, 3000, 3500, or 4000 nucleotides long.

In other embodiments the product is less than or equal to 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1250, 1500, 1750, 2000, 2250, 2500, 2750, 3000, 3500, or 4000 nucleotides long.

3. Kits

Disclosed herein are kits that are drawn to reagents that can be used in practicing the methods disclosed herein. The kits can include any reagent or combination of reagent discussed herein or that would be understood to be required or beneficial in the practice of the disclosed methods. For example, the kits could include primers to perform the amplification reactions discussed in certain embodiments of the methods, as well as the buffers and enzymes required to use the primers as intended. For example, disclosed is a kit for assessing a subject's risk for acquiring prostate cancer, comprising the oligonucleotides set forth in SEQ ID Nos:

1 and 2.

C. Methods of making the compositions

The compositions disclosed herein and the compositions necessary to perform the disclosed methods can be made using any method known to those of skill in the art for that particular reagent or compound unless otherwise specifically noted. For example, the nucleic acids, such as, the oligonucleotides to be used as primers can be made using standard chemical synthesis methods or can be produced using enzymatic methods or any other known method. Such methods can range from standard enzymatic digestion followed by nucleotide fragment isolation (see for example, Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd Edition (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989) Chapters 5, 6) to purely synthetic methods, for example, by the cyanoethyl phosphoramidite method using a Milligen or Beckman System 1Plus DNA synthesizer (for example, Model 8700 automated synthesizer of Milligen-Biosearch, Burlington, MA or ABI Model 380B). Synthetic methods useful for making oligonucleotides are also described by Ikuta *et al.*, *Ann. Rev. Biochem.* 53:323-356 (1984), (phosphotriester and phosphite-triester methods), and Narang *et al.*, *Methods Enzymol.*, 65:610-620 (1980), (phosphotriester method). Protein nucleic acid molecules can be made using known methods such as those described by Nielsen *et al.*, *Bioconjug. Chem.* 5:3-7 (1994).

D. Methods of using the compositions

It is understood that any variation or implementation of the compositions discussed herein is understood to be a particular embodiment which can be used and is described as being used in the disclosed methods. Disclosed are methods for assessing the risk that a subject may acquire prostate cancer. Methods are also disclosed for assessing the clinical significance of the prostate cancer that a subject will get or already has. These methods involve assaying the length of regions of genes that have been shown herein to be related to the onset and severity of prostate cancer. The methods involve assaying regions of both the AIB1 gene and the AR gene either individually or collectively. There are particular regions of the AIB1

gene and the AR gene which have been shown to be related to the onset and severity of prostate cancer and these regions are found in all forms and variants of these genes. In one embodiment of the disclosed methods the length of the CAG/CAA repeat in either the AIB1 gene or AR gene of a subject, discussed above, is determined. This length can then be compared to known lengths, which have been shown to be correlated with the onset and severity of prostate cancer. For example, in the AIB1 gene, the CAG/CAA region determined to be greater than, less than, or equal to 29 repeats. When the number of repeats present in the repeat region of the subject is less than or more than 29 at least one allele of the subject's AIB1 gene, this person is determined to have an increased risk of developing prostate cancer. It is understood that since the AIB1 gene is present on chromosome 20, subjects will have two alleles of this gene. When a single allele has a length different than 29 the subject has an increased risk of acquiring prostate cancer. If, however, the subject has two alleles of the AIB1 gene that both have lengths in the CAG/CAA repeat region different than 29, the subject has a higher risk of acquiring prostate cancer than if only one allele has a CAG/CAA length different than 29. Thus, if the alleles of the AIB1 gene are assayed individually, if the first allele assayed has a CAG/CAA repeat length different than 29, a certain amount of information as to the person's susceptibility will be gained, however, if the allele contains a CAG/CAA repeat equal to 29, the second allele must be looked at to get meaningful information as to the person's likelihood of acquiring prostate cancer. Furthermore, more information will be gained by assaying the second allele even if the first allele has a CAG/CAA length less than or greater than 29 because the second allele may also have a CAG/CAA length different 29 and this provides more information about the subject's prostate cancer susceptibility. Thus, it is preferred that both alleles of the AIB1 gene are assayed when performing the methods herein.

The CAG/CAA region of the subject's AR gene can also be assayed. However, in the AR length the length of the subject's CAG/CAA region in the AR gene is compared to 23 repeats. If there are more or less than 23 repeats the subject has an increased risk of prostate cancer. Furthermore, the risk of prostate cancer in a

subject is directly correlated with the length of the subject's CAG/CAA repeat. The greater the deviation of the number of CAG/CAA repeats in a subjects DNA from 23 repeats the greater the likelihood that the person will acquire prostate cancer.

1. Methods assaying the CAG/CAA repeat of the AIB1 gene

Methods are disclosed where only the length of CAG/CAA repeat in a subject's AIB1 gene is assayed. Disclosed are methods for assessing the risk of developing prostate cancer in a human by analyzing the AIB1 gene of the subject.

Disclosed are methods for assessing the risk of prostate cancer in a human subject comprising determining the length of the contiguous CAG or CAA repeats in both AIB1 gene alleles of the subject and assessing whether the length of the CAG or CAA repeats is less than, equal to, or greater than 29 repeats, a length less than or greater than 29 repeats in both alleles indicating an increased risk of prostate cancer in the subject.

Also disclosed are methods where determining the length of the repeats comprises amplifying a region of both AIB1 gene alleles comprising the contiguous CAG or CAA repeat. Any type of amplification method can be employed to amplify the target regions. Preferred methods include PCR and direct sequencing of the target repeat regions.

When two repeat regions of two alleles are amplified two products will be produced. These products can then be assayed for their length, by for example, gel chromotagraphy, HPLC, or capillary gel electrophoresis. Thus, disclosed are methods that produce two PCR products and methods where the PCR products are analyzed by chromatography, including but not limited to gel electrophoresis.

In certain embodiments the sequence of the repeat region can be determined by, for example, direct sequencing of the repeat region. In some embodiments the PCT product or other DNA amplification product, can itself be sequenced.

In certain embodiments of the disclosed methods, the repeat region is assayed or amplified by targeting primers or probes to the region or the areas of the target gene surrounding the region. These primers, for example, can be used to

amplify the region by for example, PCR. It is preferred that the method utilize primers as they are discussed herein. For example, the PCR product can be produced using a first AIB1 primer that selectively hybridizes with the complement of sequence 5' to the repeat region and a second AIB1 primer that selectively hybridizes with sequence 3' to the repeat region.

In certain embodiments, the PCR product can be produced using a first AIB1 primer that selectively hybridizes with the sequence set forth in SEQ ID NO:3 and a second AIB1 primer that selectively hybridizes with the sequence set forth in SEQ ID NO:4. In still other embodiments of the disclosed methods, the first AIB1 primer has the sequence set forth in SEQ ID NO:1 and the second AIB1 primer has the sequence set forth in SEQ ID NO:2.

In certain embodiments determining how many CAG or CAA repeats there are comprises sequencing the CAG or CAA repeats.

2. Methods assaying the CAG/CAA repeat of the Androgen receptor gene

Just as with the AIB1 gene, there are repeat regions within the androgen receptor gene that can be assayed and used to assess a person's risk for acquiring prostate cancer. There are two repeat regions that can be assayed within the androgen receptor gene. The CAG/CAA repeat region within the AR gene can be assayed as discussed for the CAG/CAA repeat region of the AIB1 gene. There is a key difference, however, and that is the length of the subject's CAG/CAA repeat region in the AR gene is compared to 23 repeats, not 29 repeats. The predictability of acquiring prostate cancer arises from a comparison to 23 CAG/CAA repeats. Furthermore, the CAG/CAA repeat with in the AR gene is predictive based on single repeat changes. Thus, the likelihood of acquiring prostate cancer increases for each difference in length from 23 repeats. Thus, for example, a subject having 15 repeats is more likely to acquire prostate cancer or have a clinically significant prostate cancer than a person having 16 repeats and a subject having 16 repeats is more likely to acquire prostate cancer or have a clinically significant prostate cancer than a person having 17 repeats and a subject having 17 repeats is more likely to acquire

prostate cancer or have a clinically significant prostate cancer than a person having 16 repeats and so forth. This relationship holds for all lengths of CAG/CAA repeats in the AR gene.

The AR gene also has a GGN repeat region whose length is related to the susceptibility of prostate cancer. The methods and reagents disclosed for the CAG/CAA repeats of the AR gene apply to the GGN repeat within the AR gene and are all specifically contemplated herein.

3. Methods assaying the CAG/CAA repeat of the Androgen receptor gene and the AIB1 gene

Methods are also disclosed where both the AIB1 gene and the AR gene are assayed to arrive at a likelihood that a subject will acquire prostate cancer. It is understood that all of the conditions and permutations of performing the methods that involve assaying either AIB1 or AR alone can also be used and applied when both AIB1 and AR are assayed.

Disclosed are methods for assessing the risk of developing prostate cancer in a human by analyzing the AIB1 gene and the AR gene of the subject together.

For example, disclosed are methods for assessing the risk of prostate cancer in a human subject comprising determining the length of the contiguous CAG or CAA repeats in the AIB1 gene alleles of the subject and assessing whether the length of the CAG or CAA repeats in each allele is less than, equal to, or greater than 29 repeats, and determining the length of the contiguous CAG or CAA repeats in the androgen receptor gene of the subject and assessing whether the length of the CAG or CAA repeats is less than, equal to, or greater than 23 repeats, a length in at least one allele less than or greater than 29 repeats in the AIB1 gene and less than or greater than 23 repeats in the androgen receptor gene indicating an increased risk of prostate cancer in the subject.

Also disclosed are methods wherein determining the length of the contiguous CAG or CAA repeats in the AIB1 gene alleles comprises amplifying a region of the AIB1 gene alleles comprising the CAG or CAA repeats and wherein determining the

length of the CAG or CAA repeats in the androgen receptor gene comprises amplifying a region of the androgen receptor gene comprising the CAG or CAA repeats.

Furthermore, methods are disclosed wherein the amplification of the regions of the AIB1 gene and the androgen receptor gene is by PCR that produces a first and a second AIB1 PCR product and an androgen receptor PCR product.

Disclosed are methods further comprising analyzing the PCR products by chromatography and methods wherein the chromatography is gel electrophoresis. The sequence of the PCR products can be determined.

The methods involving amplifying repeat regions in both AIB1 and the AR genes can for example involve primers. It is understood that primers specific for each repeat region would be used. For example, the AIB1 PCR product is produced using a first AIB1 primer that selectively hybridizes with the sequence set forth in SEQ ID NO:3 and a second AIB1 primer that selectively hybridizes with the sequence set forth in SEQ ID NO:4. Additional methods are wherein the first AIB1 primer has the sequence set forth in SEQ ID NO:1 and the second AIB1 primer has the sequence set forth in SEQ ID NO:2.

Or when assaying both genes the androgen receptor PCR product is produced using a first androgen receptor CAG primer that selectively hybridizes with the sequence set forth in SEQ ID NO:9 and a second androgen receptor CAG primer that selectively hybridizes with the sequence set forth in SEQ ID NO:10. In some embodiments the first androgen receptor CAG primer has the sequence set forth in SEQ ID NO:7 and the second androgen receptor CAG primer has the sequence set forth in SEQ ID NO:8.

Or when both the AIB1 and the AR genes are assayed, the AIB1 PCR product is produced using a first AIB1 primer that selectively hybridizes with the sequence set forth in SEQ ID NO:3 and a second AIB1 primer that selectively hybridizes with the sequence set forth in SEQ ID NO:4. and wherein the androgen receptor PCR product is produced using a first androgen receptor CAG primer that

selectively hybridizes with the sequence set forth in SEQ ID NO:9 and a second androgen receptor CAG primer that selectively hybridizes with the sequence set forth in SEQ ID NO:10.

In certain embodiments the first AIB1 primer has the sequence set forth in SEQ ID NO:1 and the second AIB1 primer has the sequence set forth in SEQ ID NO:2 and wherein the first androgen receptor CAG primer has the sequence set forth in SEQ ID NO:7 and the second androgen receptor CAG primer has the sequence set forth in SEQ ID NO:8.

In some embodiments the methods are performed wherein more than 29 contiguous CAG or CAA repeats in at least one allele of the AIB1 gene of the person and more than 23 contiguous CAG or CAA repeats in the androgen receptor gene of the person indicates an increased risk of prostate cancer in the subject.

In other embodiments the methods are performed wherein more than 29 contiguous CAG or CAA repeats in at least one allele of the AIB1 gene of the person and less than 23 contiguous CAG or CAA repeats in the androgen receptor gene of the person indicates an increased risk of prostate cancer in the subject.

In some embodiments the methods are performed wherein less than 29 contiguous CAG or CAA repeats in at least one allele of the AIB1 gene of the person and more than 23 contiguous CAG or CAA repeats in the androgen receptor gene of the person indicates an increased risk of prostate cancer in the subject.

In other embodiments the methods are performed wherein less than 29 contiguous CAG or CAA repeats in at least one allele of the AIB1 gene of the person and less than 23 contiguous CAG or CAA repeats in the androgen receptor gene of the person indicates an increased risk of prostate cancer in the subject.

In other embodiments the methods are performed wherein more or less than 29 contiguous CAG or CAA repeats in both alleles of the AIB1 gene of the person and more than 23 contiguous CAG or CAA repeats in the androgen receptor gene of the person indicates an increased risk of prostate cancer in the subject.

In other embodiments the methods are performed wherein more or less than 29 contiguous CAG or CAA repeats in both alleles of the AIB1 gene of the person and less than 23 contiguous CAG or CAA repeats in the androgen receptor gene of the person indicates an increased risk of prostate cancer in the subject.

In other embodiments the methods are performed wherein more than 29 contiguous CAG or CAA repeats in both alleles of the AIB1 gene of the person and more than 23 contiguous CAG or CAA repeats in the androgen receptor gene of the person indicates an increased risk of prostate cancer in the subject.

In other embodiments the methods are performed wherein less than 29 contiguous CAG or CAA repeats in both alleles of the AIB1 gene of the person and less than 23 contiguous CAG or CAA repeats in the androgen receptor gene of the person indicates an increased risk of prostate cancer in the subject.

Also disclosed are methods for assessing the risk of prostate cancer in a human subject comprising determining the length of the contiguous CAG or CAA repeats in the AIB1 gene alleles of the subject and assessing whether the length of the CAG or CAA repeats in each allele is less than, equal to, or greater than 29 repeats, and determining the length of the contiguous GGN repeats in the androgen receptor gene of the subject and assessing whether the length of the GGN repeats is less than, equal to, or greater than 23 repeats, a length in at least one allele less than or greater than 29 repeats in the AIB1 gene and less than or greater than 23 repeats in the androgen receptor gene indicating an increased risk of prostate cancer in the subject, wherein N is either T, G, or C.

Disclosed are methods wherein determining the length of the contiguous CAG or CAA repeats in the AIB1 gene alleles comprises amplifying a region of the AIB1 gene alleles comprising the CAG or CAA repeats and wherein determining the length of the contiguous GGN repeats in the androgen receptor gene comprises amplifying a region of the androgen receptor gene comprising the GGN repeats.

Further methods are disclosed wherein the amplification of the regions of the AIB1 gene and the androgen receptor gene is by PCR that produces a first and a

second AIB1 PCR product and an androgen receptor PCR product.

In other methods the amplification further comprises analyzing the PCR products by chromatography.

The amplification products including the PCR products can be analyzed by gel electrophoresis.

In other embodiments the sequence of the PCR products is determined.

Disclosed are amplification methods wherein the AIB1 PCR product is produced using a first AIB1 primer that selectively hybridizes with the sequence set forth in SEQ ID NO:3 and a second AIB1 primer that selectively hybridizes with the sequence set forth in SEQ ID NO:4.

Also disclosed are amplification methods wherein the first AIB1 primer has the sequence set forth in SEQ ID NO:1 and the second AIB1 primer has the sequence set forth in SEQ ID NO:2.

Disclosed are methods wherein the androgen receptor PCR product is produced using a first androgen receptor GGN primer that selectively hybridizes with the sequence set forth in SEQ ID NO:13 and a second androgen receptor GGN primer that selectively hybridizes with the sequence set forth in SEQ ID NO:14.

Also disclosed are methods wherein the first androgen receptor GGN primer has the sequence set forth in SEQ ID NO:11 and the second androgen receptor GGN primer has the sequence set forth in SEQ ID NO:12.

Further disclosed are methods wherein the AIB1 PCR product is produced using a first AIB1 primer that selectively hybridizes with the sequence set forth in SEQ ID NO:3 and a second AIB1 primer that selectively hybridizes with the sequence set forth in SEQ ID NO:4. and wherein the androgen receptor PCR product is produced using a first androgen receptor GGN primer that selectively hybridizes with the sequence set forth in SEQ ID NO:13 and a second androgen receptor GGN primer that selectively hybridizes with the sequence set forth in SEQ ID NO:14.

Also disclosed are methods wherein the first AIB1 primer has the sequence

set forth in SEQ ID NO:1 and the second AIB1 primer has the sequence set forth in SEQ ID NO:2 and wherein the first androgen receptor GGN primer has the sequence set forth in SEQ ID NO:11 and the second androgen receptor GGN primer has the sequence set forth in SEQ ID NO:12.

In some embodiments the methods are performed wherein more than 29 contiguous CAG or CAA repeats in at least one allele of the AIB1 gene of the person and more than 23 contiguous GGN repeats in the androgen receptor gene of the person indicates an increased risk of prostate cancer in the subject.

In other embodiments the methods are performed wherein more than 29 contiguous CAG or CAA repeats in at least one allele of the AIB1 gene of the person and less than 23 contiguous GGN repeats in the androgen receptor gene of the person indicates an increased risk of prostate cancer in the subject.

In some embodiments the methods are performed wherein less than 29 contiguous CAG or CAA repeats in at least one allele of the AIB1 gene of the person and more than 23 contiguous GGN repeats in the androgen receptor gene of the person indicates an increased risk of prostate cancer in the subject.

In other embodiments the methods are performed wherein less than 29 contiguous CAG or CAA repeats in at least one allele of the AIB1 gene of the person and less than 23 contiguous GGN repeats in the androgen receptor gene of the person indicates an increased risk of prostate cancer in the subject.

In some embodiments the methods are performed wherein more or less than 29 contiguous CAG or CAA repeats in both alleles of the AIB1 gene of the person and more than 23 contiguous GGN repeats in the androgen receptor gene of the person indicates an increased risk of prostate cancer in the subject.

In other embodiments the methods are performed wherein more or less than 29 contiguous CAG or CAA repeats in both alleles of the AIB1 gene of the person and less than 23 contiguous GGN repeats in the androgen receptor gene of the person indicates an increased risk of prostate cancer in the subject.

In other embodiments the methods are performed wherein more than 29 contiguous CAG or CAA repeats in both alleles of the AIB1 gene of the person and more than 23 contiguous GGN repeats in the androgen receptor gene of the person indicates an increased risk of prostate cancer in the subject.

In other embodiments the methods are performed wherein less than 29 contiguous CAG or CAA repeats in both alleles of the AIB1 gene of the person and less than 23 contiguous GGN repeats in the androgen receptor gene of the person indicates an increased risk of prostate cancer in the subject.

Also disclosed are methods for assessing the risk of prostate cancer in a human subject comprising determining the length of the contiguous CAG or CAA repeats in an AIB1 gene of the subject and assessing whether the length of the CAG repeats is less than, equal to, or greater than 29 repeats, a length less than or greater than 29 repeats indicating an increased risk of prostate cancer in the subject.

Methods for assessing the risk of prostate cancer in a human subject comprising determining the length of the contiguous CAG or CAA repeats in an AIB1 gene of the subject and assessing whether the length of the CAG or CAA repeats is less than, equal to, or greater than 29 repeats, and determining the length of the contiguous CAG or CAA repeats in the androgen receptor gene of the subject and assessing whether the length of the CAG or CAA repeats is less than, equal to, or greater than 23 repeats, a length less than or greater than 29 repeats in the AIB1 allele and less than or greater than 23 repeats in the androgen receptor gene indicating an increased risk of prostate cancer in the subject are also disclosed.

Disclosed are methods for assessing the risk of prostate cancer in a human subject comprising determining the length of the contiguous CAG or CAA repeats in an AIB1 gene of the subject and assessing whether the length of the CAG or CAA repeats is less than, equal to, or greater than 29 repeats, and determining the length of the contiguous GGN repeats in the androgen receptor gene of the subject and assessing whether the length of the GGN repeats is less than, equal to, or greater than 23 repeats, a length less than or greater than 29 repeats in the AIB1 allele and

less than or greater than 23 repeats in the androgen receptor gene indicating an increased risk of prostate cancer in the subject, wherein N is either T, G, or C.

Throughout this application, various publications are referenced. The disclosures of these publications in their entireties are hereby incorporated by reference into this application in order to more fully describe the state of the art to which this invention pertains. The references disclosed are also individually and specifically incorporated by reference herein for the material contained in them that is discussed in the sentence in which the reference is relied upon.

It will be apparent to those skilled in the art that various modifications and variations can be made in the present invention without departing from the scope or spirit of the invention. Other embodiments of the invention will be apparent to those skilled in the art from consideration of the specification and practice of the invention disclosed herein. It is intended that the specification and examples be considered as exemplary only, with a true scope and spirit of the invention being indicated by the following claims.

E. Examples

The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how the compounds, compositions, articles, devices and/or methods claimed herein are made and evaluated, and are intended to be purely exemplary of the invention and are not intended to limit the scope of what the inventors regard as their invention. Efforts have been made to ensure accuracy with respect to numbers (e.g., amounts, temperature, etc.), but some errors and deviations should be accounted for. Unless indicated otherwise, parts are parts by weight, temperature is in °C or is at ambient temperature, and pressure is at or near atmospheric.

1. Example 1

Because AR coactivators enhance transactivation of AR, the relationship of a CAG/CAA repeat length polymorphism in the *AIB1/SRC-3* gene (amplified in breast cancer gene 1, a steroid receptor coactivator and an AR coactivator) with prostate cancer risk in a multidisplinary population-based case-control study in China was

evaluated. Genomic DNA of 189 prostate cancer patients and 301 healthy controls was used for the PCR-based assay. The *AIB1/SRC-3* CAG/CAA repeat length for all alleles ranged from 24 to 32, with the most common repeat length being 29. Homozygous 29/29 and heterozygous 28/29 were the most common genotypes, with 44% and 30% of the controls harboring these genotypes, respectively. Relative to subjects homozygous for ≥29 CAG repeats (≥29/≥29 genotype), individuals carrying one shorter allele (≥29/<29 genotype) had a 26% increased risk (OR = 1.26, 95% CI = 0.85-1.86), while those homozygous for the shorter allele (<29/<29 genotype) had a 73% excess risk (OR = 1.73, 95% CI = 0.96-3.11). The combined effect of CAG repeat lengths in the *AR* and *AIB1/SRC*-3 genes was also evaluated. Relative to men with both the ≥29/≥29 genotype of the *AIB1/SRC-3* gene and a long CAG repeat length (≥23) in the *AR* gene, those with both the <29/<29 *AIB1/SRC-3* genotype and a short CAG repeat length in the *AR* gene (<23) had a 2.9-fold risk (OR = 2.86, 95% CI = 1.28-6.40). A similar was seen for the combined effects of the *AIB1/SRC-3* marker with the GGN marker of the *AR* gene. Together, these data indicate that the CAG/CAA repeat length in the *AIB1/SRC-3* gene is associated with prostate cancer risk in Chinese men and that the combination of CAG/CAA repeat lengths in both the *AIB1/SRC-3* and *AR* genes provide a useful marker for clinically significant prostate cancer.

a) Materials and methods

(1) Study subjects

Details of the study have been described previously (9,16-18 AIB1) which are herein are incorporated by reference for at least material related to the epidemiological study . Cases of primary prostate cancer (ICD9 185) newly diagnosed between 1993 and 1995 were identified through a rapid reporting system established between the Shanghai Cancer Institute and 16 collaborating hospitals in urban Shanghai. Cases were permanent residents in 10 urban districts of Shanghai (henceforth referred to as Shanghai) who had no history of other cancer. Contrary to many Western countries, prostate cancer screening is not widespread in China; therefore cases in this study were clinically significant prostate cancers who

presented with symptoms.

Based on the personal registry cards of all adults over age 18 residing in urban Shanghai (maintained at the Shanghai Resident Registry), male population controls were selected randomly from the 6.5 million permanent residents of Shanghai and frequency-matched to the expected age distribution (5-year category) of prostate cancer cases. Included controls were negative for prostate cancer based on digital rectal exam and trans-rectal ultrasound.

Information on potential risk factors was elicited through an in-person interview by trained interviewers using a structured questionnaire. The interview included information on demographic characteristics; dietary and smoking history; consumption of alcohol and other beverages; medical history; family history of cancer; physical activity; body size; and sexual behavior. Of the 268 eligible cases (95% of the cases diagnosed in Shanghai during the study period), 243 (91%) were interviewed. After a consensus review by both the Chinese and American pathologists, four cases were classified as having benign prostatic hyperplasia and excluded from the study. Of the 495 eligible controls, 472 (95%) were interviewed. Most non-responses were due to refusal.

b) Blood collection and DNA extraction

Two hundred cases (82% of those interviewed) and 330 controls (70%) provided 20 ml of fasting blood for the study. The blood samples were processed at a central laboratory in Shanghai. The buffy coat samples were first stored at -70°C and then shipped to the U.S. in dry ice for DNA extraction at the American Type Culture Collection (Manassas, VA), using a standard DNA extraction protocol. Quality control procedures showed no evidence of contamination, and DNA purity and length were satisfactory. After DNA extraction, 190 cases and 305 controls had sufficient DNA for genotyping. DNA samples were arranged in case-control pairs/triplets to minimize day-to-day laboratory variation, and laboratory personnel were masked to case-control status.

c) Genotyping

The *AIB1/SRC-3* gene. The polyglutamine region of the AIB1/SRC-3 protein is encoded by two glutamine codons in the *AIB1/SRC-3* gene on chromosome 20 (GenBank accession number AF012108): CAG and CAA. The usual sense codon sequence of the polyglutamine stretch is $(CAG)_n$ CAA $(CAG)_n$ $(CAA\ CAG)_4$ CAG CAA $(CAG)_2$ CAA) SEQ ID NO:5. The two variable-length tracks of CAG repeats ($(CAG)_n$) usually contain six repeats between nucleotides 3930 and 3947, and nine repeats between nucleotides 3951 and 3977, for a total repeat length of 29 (19 AIB1) (which is herein incorporated by reference for material related to the CAG repeat.) This polymorphism has previously been described by Shirazi et al., though while Shirazi et al. scored genotypes of this marker using only the two variable $(CAG)_n$ stretches (19 AIB1), herein the data was scored as the total number of continuous CAG and CAA triplets in the entire polyglutamine region of the *AIB1/SRC-3* gene, as has been done more recently (20,21 AIB1) (both of which are herein incorporated by reference for material related to the CAG and CAA repeats).

The number of CAG/CAA repeats in the polyglutamine stretch of the *AIB1/SRC-3* gene was determined by amplifying the gene's C-terminal polyglutamine region in each sample using custom flanking primers (5'-TCATCACCTCCGACAACAGAGG-3' and 5' (SEQ ID NO:1)-TATGGAAACTGTTGCGGAGGAG-3' (SEQ ID NO:2) and the Advantage 2 Polymerase System (Clontech). The number of CAG/CAA repeats was determined by electrophoresis of the PCR products on an acrylamide gel and comparison to molecular weight standards. For confirmation, PCR products from selected samples were subsequently purified, using the PCR Product Purification Kit (Qiagen), and sequenced directly, using the Big Dye Terminator Cycle Sequencing Ready Reaction Kit (PE Biosystems).

The *AR* gene. Genotypes of both the CAG (polyglutamine) and GGN (polyglycine) repeat length polymorphisms in exon 1 of the *AR* gene (located on the X chromosome) were determined as described previously (22 AIB1) (which is

herein incorporated by reference for material related to the AR gene). Briefly, two sets of oligonucleotide primers flanking each of the two polymorphic regions for use in DNA amplification and direct sequencing were designed. For the polyglutamine stretch, the number of continuous CAG or CAA triplets was counted directly, while for the polyglycine stretch, the number of continuous GGN repeats (where N represents T, G, or C) was counted directly

Quality control. Because the PCR procedure is prone to contamination, a negative water-blank control was always included in each batch of the PCR reactions (usually 9-18 DNA samples plus one negative control). If the negative control was shown to be positive, the assay was repeated for the entire batch. Twenty-four split samples from a single individual were spaced at intervals among the study samples to assess the reproducibility of genotyping. For the *AIB1/SRC-3* gene, all of the 24 split samples had a CAG/CAA repeat length of 29 in both alleles. Of the 21 split samples with *AR* CAG results, 19 (90%) had the same repeat number (repeat length=23); one had one more, and one had one less repeat. Of the 20 samples with *AR* GGN results, 19 (95%) had the same repeat number (repeat length=23) and one had one less repeat.

Statistical Analysis, Unconditional logistic regression models were used to derive odds ratios (ORs) and corresponding 95% confidence intervals (CIs) to estimate the prostate cancer risks associated with *AIB1/SRC-3* genotypes (23 AIB1) (which is herein incorporated by reference for material related to statistical analysis). The distribution of the number of the CAG/CAA repeat lengths among controls were used to derive the median cutoffs used to calculate the ORs. Because the *AIB1/SRC-3* gene is located on chromosome 20, each individual carries two alleles. In contrast, since the *AR* gene is located on the X chromosome, there is only one allele for each individual. *AR* CAG and the *AR* GGN polymorphisms in this Chinese population (9 AIB1), subjects in the current *AIB1/SRC-3* analysis were grouped by <23 versus ≥ 23 repeats in analyses stratified by each of the two *AR* gene polymorphisms. The level of significance for all results reported herein is 0.05.

d) Results

Age at diagnosis ranged from 50 to 94 (median 73) for cancer cases. Due to the lack of widespread prostate cancer screening in China, cases in this study were mostly men with clinically significant prostate cancer. Accordingly, about two-thirds of the cases were diagnosed as having advanced (regional/remote stages) cancer, and most tumors were moderately or poorly differentiated. Most cases were symptomatic at diagnosis, and 77% had serum prostatic specific antigen levels greater than 10 ng/ml (median 87 ng/ml). Compared to population controls, cases had significantly higher caloric intake; had significantly larger waist-to-hip ratios; and were somewhat less likely to be married, have attended college, or be smokers or drinkers, though not significantly so (data not shown).

The distribution of the alleles and genotypes of the *AIB1/SRC-3* gene CAG/CAA repeat length marker by case-control status is shown in Table 1. Among controls, the CAG/CAA repeat length for all alleles ranged from 24 to 32, with 29, 28, and 26 being the most common repeat lengths (65.6, 23.5, and 6.5%, respectively). Eighty-eight percent of the controls had at least one 29 allele. Homozygous 29/29 and heterozygous 28/29 were the most common genotypes, with 44% and 30% of the controls harboring these genotypes, respectively. The observed genotype frequencies were in close agreement with those predicted from the allele frequencies under Hardy-Weinberg equilibrium.

Table 1. **Frequencies of *AIB1/SRC-3* CAG/CAA repeat length alleles and genotypes in prostate cancer cases and controls, China**

	Cases		Controls	
AIB1/SRC-3 allele	n	%	n	%
24	0	0.0	1	0.2
26	22	5.8	39	6.5
27	6	1.6	7	1.2
28	114	30.0	140	23.5
29	224	58.9	391	65.6
30	11	2.9	13	2.2
31	1	0.3	5	0.8

32	2	0.5	0	0.0
Total	380		596	
AIB1/SRC-3 genotype	n	%	n	%
24/29	0	0.0	1	0.3
26/26	0	0.0	1	0.3
26/27	0	0.0	1	0.3
26/28	5	2.6	6	2.0
26/29	17	8.9	26	8.7
26/30	0	0.0	2	0.7
26/31	0	0.0	2	0.7
27/28	0	0.0	1	0.3
27/29	5	2.6	5	1.7
27/30	1	0.5	0	0.0
28/28	23	12.1	21	7.0
28/29	59	31.1	88	29.5
28/30	3	1.6	1	0.3
28/31	1	0.5	2	0.7
29/29	67	35.3	130	43.6
29/30	7	3.7	10	3.4
29/31	0	0.0	1	0.3
29/32	2	1.1	0	0.0
Total	190		298	

Relative to men homozygous for 29 CAG/CAA repeats in the *AIB1/SRC-3* gene (29/29 genotype), subjects homozygous for the 28 allele had a significant risk increase (OR = 2.12, 95% CI = 1.09-4.12, Table 2). Subjects with the 28/29 genotype had a non-significant increased risk relative to the 29/29 genotype (OR = 1.30, 95% CI = 0.83-2.03), as did men with one 29 allele and one 24, 26, or 27 allele (OR = 1.33, 95% CI = 0.72-2.47) and men with one 29 allele and one 30, 31, or 32 allele (OR = 1.58, 95% CI = 0.62-4.01).

Table 2. **Age-adjusted ORs for prostate cancer in relation to CAG/CAA repeat lengths in the *AIB1/SRC-3* gene, China**

Allele 1	Allele 2	Cases	Control	OR[a]	95% CI
29	29	67	130	1.00	-
28	29	59	88	1.30	0.83-2.03
24,26,27	29	22	32	1.33	0.72-2.47
30,31,32	29	9	11	1.58	0.62-4.01
28	28	23	21	2.12	1.09-4.12
26,27,28	30,31	5	7	1.38	0.42-4.52

26,27	26,27,28	5	9	1.08	0.35-3.35
≥29	≥29	76	141	1.00	-
≥29	<29	86	127	1.26	0.85-1.86
<29	<29	28	30	1.73	0.96-3.11

Based on the median CAG/CAA repeat length of 29, the various *AIB1/SRC-3* alleles were collapsed into those with 29 or more repeats (≥29) and those with less than 29 repeats (<29). Relative to men with the ≥29/≥29 genotype, men with one long and one short allele (≥29/<29 genotype) had a moderate but non-significant risk elevation (OR = 1.26, 95% CI = 0.85-1.86). Those with two short alleles (<29/<29 genotype) had a non-significant 73% increased risk (OR = 1.73, 95% CI = 0.96-3.11) relative to men with two long alleles (≥29/≥29 genotype).

Risks of prostate cancer associated with various repeat lengths in both the *AIB1/SRC-3* gene CAG/CAA polymorphism as well as the two polymorphisms of the *AR* gene are shown in Table 3. Relative to men both homozygous for the long CAG/CAA allele (≥29/≥29 genotype) of the *AIB1/SRC-3* gene and having a long *AR* CAG repeat length (≥23 repeats), men both homozygous for the short *AIB1/SRC-3* CAG/CAA allele (<29/<29) and having a short *AR* CAG repeat length (<23) had a significant 2.9-fold risk (OR=2.86, 95% = 1.28-6.40). Similarly, those men both homozygous for the short *AIB1/SRC-3* CAG/CAA allele and having a short GGN repeat length (<23) in the *AR* gene had a non-significant 2.3-fold risk (OR=2.29, 95% CI = 0.73-7.15) relative to men both homozygous for long *AIB1/SRC-3* CAG/CAA allele (≥29/≥29 genotype) and having a long *AR* GGN repeat length (≥23 repeats).

Table 3. **ORs [a] for prostate cancer in relation to *AIB/SRC-3* CAG/CAA and *AR* CAG or GGN repeat lengths, China**

AR polymorphisms	CAG/CAA repeat length genotypes of the *AIB1/SRC-3* gene								
	≥29/≥29			≥29/<29			<29/<29		
	n1/n2[b]	OR	95% CI	n1/n2[b]	OR	95% CI	n1/n2[b]	OR	95% CI
CAG repeat length[c]									
≥23	30/68	1.00	-	35/69	1.15	0.63-2.08	9/14	1.46	0.57-3.75
<23	45/71	1.43	0.81-2.54	51/58	1.99	1.12-3.53	19/15	2.86	1.28-6.40
GGN repeat length[c]									
≥23	57/112	1.00	-	72/99	1.42	0.92-2.21	19/23	1.62	0.81-3.22
<23	17/26	1.28	0.64-2.56	14/24	1.15	0.55-2.40	7/6	2.29	0.73-7.15

[a] Adjusted for age.
[b] n1=number of cases, n2=number of controls.
[c] Median number of repeats among controls was used for the cutoff.

There was no correlation between the repeat lengths in the *AR* and *AIB1/SRC-3* genes. In addition, the number of CAG/CAA repeats in the *AIB1/SRC-3* gene did not correlate with education, body mass index, waist-to-hip ratio, total caloric intake, serum levels of sex hormones (testosterone; DHT; 5α-androstane-3α,17β-diol glucuronide; and estradiol), and sex hormone binding globulin. These variables therefore were not included in the logistic model for adjustment. In addition, odds ratios were materially unchanged when the analysis was stratified by clinical stage (localized versus advanced stage disease, data not shown).

Results from this population-based case-control study in China indicate that men with fewer than 29 CAG/CAA repeats in the *AIB1/SRC-3* alleles have an increased risk of clinically significant prostate cancer. Furthermore, our results suggest that this effect, though independent of *AR* genotypes, is more pronounced among men with a smaller number of *AR* CAG repeats.

The observed association with CAG/CAA repeat length in the *AIB1/SRC-3* gene is biologically plausible. Data from transient transfection studies show that the AIB1/SRC-3 coactivator enhances AR transcriptional activity in the presence of

DHT (12 AIB1), suggesting that the AIB1/SRC-3 coactivator, in conjunction with AR, may increase androgenic activity within the prostate gland. Amplification of the *AIB1/SRC-3* gene has been implicated in the etiology of several other hormone-dependent cancers as well, including breast and ovarian cancers (24 AIB1). Furthermore, recent clinical data suggest that overexpression of AR in prostate tumors may contribute to hormone sensitivity and tumor progression (25 AIB1).

Racial/ethnic variation in the *AIB1/SRC-3* CAG/CAA repeat length mirrors the risk patterns of prostate cancer in high- and low-risk populations (19,26 AIB1), thus indirectly supporting a role of AIB1/SRC-3 in prostate cancer etiology. In a small survey of 112 African Americans, 19 Chinese, and 18 Caucasians, the allele frequency of 29 CAG/CAA repeats was 61%, 76%, and 58%, respectively.

Chinese men have a longer mean CAG repeat length in the *AR* gene than Western men, and that a shorter *AR* CAG repeat length was associated with an increased risk of prostate cancer in this low-risk population (9 AIB1). The observed association with CAG/CAA repeat length in the *AIB1/SRC-3* gene is independent of the *AR* gene: regardless of CAG repeat length in the *AR* gene, men homozygous for the shorter *AIB1/SRC-3* allele (the $<29/<29$ genotype) had a higher risk than those homozygous for the long allele ($\geq 29/\geq 29$ genotype). However, the risk associated with the homozygous $<29/<29$ *AIB1/SRC-3* genotype was more pronounced among those with the short *AR* CAG repeat length.

2. Example 2. Relationship of CAG length in the androgen receptor

The length of the polymorphic CAG trinucleotide repeat in the polyglutamine region of the androgen receptor (AR) gene is inversely correlated with the transactivation function of the AR. A population-based case-control assay in China addressed CAG and other polymorphisms of the AR gene and their association with clinically significant prostate cancer in this low-risk population. Genomic DNA from 190 prostate cancer patients and 304 healthy controls were used for direct sequencing to evaluate the relationship of CAG and GGN (polyglycine) repeat length in the AR gene. Relative to western men, the subjects had a longer

CAG repeat length, with a median of 23 and only 10% of the subjects having a CAG repeat length shorter than 20. Men with a CAG repeat length shorter than 23 (median length) had a 65% increased risk of prostate cancer (OR=1.65, 95% CI 1.14-2.39), compared to men with a CAG repeat length of 23 or longer.

For the GGN tract ($GGT_3GGG_1GGT_2GGC_n$), based on the sequencing results from 481 samples, it is shown that even though GGC regions in the polyglycine tract are highly variable, there are no mutations or polymorphisms in the GGT and GGG regions. Seventy two percent of the subjects had a GGN repeat length of 23, and those with a GGN repeat length shorter than 23 had a 12% increased risk of prostate cancer (95% CI 0.71-1.78), compared to those with 23 or more GGN repeats. This not only confirms that Chinese men do have a longer CAG repeat length than western men, but also represents the first population assay to show that even in a very low-risk population, a shorter CAG repeat length confers a higher risk of clinically significant prostate cancer. These results indicate that CAG repeat length can serve as a useful marker to identify a subset of individuals at higher risk of developing clinically significant prostate cancer.

a) Materials and methods

Subjects. Details of the assay have been described previously (28AR which is herein incorporate by reference for material related to the assay). Briefly, cases of primary prostate cancer (ICD9 185) newly diagnosed between 1993 and 1995 were identified through a rapid reporting system established between the Shanghai Cancer Institute (SCI) and 28 collaborating hospitals in urban Shanghai. Cases were permanent residents in 10 urban districts of Shanghai (henceforth referred to as Shanghai) who had no history of any other cancer. Of the 268 eligible cases (representing 95% of the cases diagnosed in urban Shanghai during the study period), 243 (91%) were interviewed in person by trained interviewers. Four of the cases were later classified as having benign prostatic hyperplasia and excluded from the study after a consensus review by both Chinese and U.S. pathologists.

Based on the records at the Shanghai Resident Registry, which contains personal identification cards for all adult residents over age 18 in urban Shanghai,

healthy subjects who were free of cancer were selected randomly from permanent residents of Shanghai (6.5 million), frequency-matched to the expected age distribution (5-year category) of prostate cancer cases. Of the 495 eligible controls without a history of cancer, 472 (95%) were interviewed.

Information on potential risk factors was elicited through an in-person interview by trained interviewers using a structured questionnaire. The interview included information on demographic characteristics; dietary history; smoking history; consumption of alcohol and other beverages; medical history; family history of cancer; physical activity; body size; and sexual behavior.

Blood collection and DNA extraction. Two hundred cases (84% of those interviewed) and 330 controls (70%) provided 20 ml of fasting blood for the study. The blood samples were processed within three hours of collection at a central laboratory in Shanghai and stored at -70°C. The frozen buffy coat samples (separated from 5 ml of blood) were later shipped to the U.S. on dry ice for DNA extraction at the American Type Culture Collection (Manassas, Virginia) with standard protocols. DNA purity, yield, and length were satisfactory and there was no evidence of DNA degradation or RNA contamination. Following DNA extraction, 191 cases and 304 controls had sufficient DNA for AR genotyping at the University of Rochester. DNA samples for cases and controls were grouped into pairs to minimize the effect of day-to-day laboratory variation. Laboratory personnel were blinded to the case-control status.

Molecular analysis and assessment of the CAG and GGN repeats. As part of an ongoing molecular analysis of the AR gene, genomic DNA from the 495 subjects was used to determine the usual sense codon sequence and the exact number of CAG and GGN repeats in exon 1 of the AR gene through PCR amplification and DNA sequencing. For the CAG repeat analysis, a set of oligonucleotide primers that flank the CAG repeat, 5'-GCTCTGGGACGCAA-CCTCTCT-3' (SEQ ID NO:7) and 5'-GCAGCGACTACCGCATCATCA-3' (SEQ ID NO:8), were designed for PCR amplification. A pair of nested primers, 5'-CGGG-GTAAGGGAAGTAGGTGGAAG-3' (SEQ ID NO:15), and 5'-

CTCTACGATGGGCTTGGGGAGAAC-3' (SEQ ID NO:16) was selected for DNA sequencing. For GGN analysis, the oligonucleotide primers 5'-ACCCTCAGCCGCCGCTTCCTCATC-3' (SEQ ID NO:11)and 5'-CTGGGATAGGGCACTCTGCTCAAC-3' (SEQ ID NO:12) were used for both PCR amplification and sequencing. The PCR products of the CAG and GGN repeats were amplified, using the Advantage 2 Polymerase System (Clontech) and the Advantage-GC Genomic Polymerase System (Clontech), respectively. Subsequently, these PCR products were purified, using the PCR Product Purification Kit (Qiagen), and sequenced directly, using the Big Dye Terminator Cycle Sequencing Ready Reaction Kit (PE Biosystems). All reactions were optimized to reach consistent results, using genomic DNA samples extracted from cell lines. For the polyglutamine tract ($(CAG)_nCAA$), the number of CAG triplets was counted to yield the length of CAG repeats. For the polyglycine tract ($GGT_3GGG_1GGT_2GGC_n$) (SEQ ID NO:17), the usual sense codon sequence of the GGN tract is: three GGT, one GGG, two GGT, followed by a variable number of GGC repeats. For example, a GGN repeat length of 23 in our study corresponded to a PCR fragment of 217 bp, encompassing 3 GGT, 1 GGG, 2 GGT, and 17 GGC triplets.

Because the PCR procedure is prone to contamination, a negative control (water blank) was always included in each batch of PCR reactions (usually 9-18 samples plus one negative control). The assay for one batch (9 samples) was repeated with new reagents because of an indication of minor contamination. Because exon 1 of the AR gene is GC-rich with CAG and GGN repeats, this region is difficult to amplify. Several samples had to be amplified and sequenced more than once. Overall, five (1%) of the 495 samples could not be typed for CAG repeats due to insufficient DNA or sequencing problems, while 14 (2.8%) could not be typed for GGN repeats for similar reasons. The percentages of samples that were unsuccessfully genotyped were similar in cases and controls.

Twenty-four split samples from the same individual were included as quality control samples to assess the reproducibility of genotyping. Of the 24 quality control samples, 23 and 20 were amplified and sequenced successfully for the CAG

and GGN repeats, respectively. Of the 23 samples with CAG results, 21 (91%) had the same repeat length of 23, one had 24, and one had 22. Of the 20 samples with GGN results, 19 (95%) had the same repeat length of 23 and the other had length 24.

Statistical Analysis. The mean numbers of CAG and GGN repeats were compared in cases and controls using the t test. Unconditional logistic regression models were used to estimate odds ratios (ORs) and their corresponding 95 percent confidence intervals (CIs) for prostate cancer in relation to CAG and GGN repeat lengths (29AR). Repeat lengths were examined first as continuous variables and later as categorical variables. The distributions of the number of CAG or GGN repeats among controls were used to derive the median or tertile cutoffs used to calculate ORs. In addition, the combined effects of CAG and GGN were evaluated based on the median lengths within the controls. The relationships between age, CAG and GGN repeat length, and other variables were assessed by Spearman correlation and analysis of variance.

b) Results

Selected characteristics of cases and controls are shown in Table 4. Compared to controls, cases had higher caloric intake and higher levels of education and waist-to-hip ratio and were less likely to use cigarettes or alcohol. Age at diagnosis ranged from 50 to 94 (median 73) for cancer cases. Sixty-nine cases (36%) were diagnosed as having localized cancer, and most tumors (72%) were moderately or poorly differentiated.

Table 4 Selected characteristics of prostate cancer cases and population controls, China

	Cases (n=191)		Controls (n=304)	
Characteristics	Mean	(S.D.)	Mean	(S.D.)
Age (yrs)	72.2	(7.7)	71.9	(7.3)
Total calories (Kcal/day)	2457.0	(647)	2342.0	(731)
Height (cm)	167.9	(6.0)	167.6	(5.8)
Weight (Kg)	61.3	(8.4)	61.5	(10.1)
BMI (Kg/m2)	21.8	(2.9)	21.9	(3.3)

Waist circumference (cm)	82.6	(10.4)	82.5	(10.7)
Hip circumference (cm)	90.7	(8.9)	92.6	(8.5)
Waist-to-hip ratio	0.91	(0.05)	0.89	(0.05)
% married	89.5		92.1	
% with education greater than high school	34.6		25.3	
% smokers	56.5		65.1	
% alcohol users	31.4		42.1	
Clinical stage (%)				
Localized	36.3			
Regional	30.5			
Remote	32.1			
Histologic grade (%)				
Well differentiated	7.9			
Moderately differentiated	31.0			
Poorly differentiated	41.0			
Cannot be assessed	20.1			

Because the AR gene is located on the X chromosome, only one copy of the gene is present in men. For the polyglutamine tract (CAG_nCAA), there was no variation in the CAA sequence among the 490 samples analyzed. The number of CAG repeats ranged from 10 to 34. About 65% of the study subjects had a CAG repeat length that ranged from 21 to 24, but only 1% of the subjects had a CAG length longer than 30 repeats (Table 5).

Table 5 *Distribution of number of CAG repeats in the androgen receptor gene in prostate cancer cases and controls, China*

	Cases (n=190)		Controls (n=300)	
No. of CAG repeats	N	%	N	%
10	0	0.0	1	0.3
14	0	0.0	1	0.3

15	4	2.1	2	0.7
16	3	1.6	2	0.7
17	1	0.5	3	1.0
18	7	3.7	9	3.0
19	7	3.7	12	4.0
20	13	6.8	17	5.7
21	24	12.6	32	10.7
22	57	30.0	67	22.3
23	22	11.6	46	15.3
24	21	11.1	48	16.0
25	16	8.4	19	6.3
26	9	4.7	17	5.7
27	3	1.6	15	5.0
28	2	1.1	2	0.7
29	0	0.0	3	1.0
30	0	0.0	1	0.3
31	0	0.0	2	0.7
32	0	0.0	0	0.0
33	0	0.0	1	0.3
34	1	0.5	0	0.0
Median	22		23	

Although the median number of CAG repeats in controls was only slightly larger than that in cases (23.0 vs. 22.0), there was a shift toward longer repeat length among controls (Figure 1). For CAG repeat length shorter than 23, cases had higher percentages than controls in 6 of the 10 categories. However, for CAG repeat length longer than 22, controls had higher percentage than cases in 8 of the 12 categories. Age at diagnosis and stage of cancer were not related to CAG repeat length, with similar distribution and average number of CAG and GGN repeat lengths in various age categories and three clinical stages.

For the polyglycine tract ($GGT_3GGG_1GGT_2GGC_n$) (SEQ ID NO:17), there

was no variation in the codon usage or the number of GGT and GGG trinucleotides in all of the 481 samples analyzed, although the number of GGC repeats was highly variable. The pattern was always three GGT, one GGG, two GGT, followed by a variable number of GGC. The number of GGN repeats among study subjects ranged from 15 to 27 (the number of GGC repeats thus ranged from 9 to 21) (Table 6). About 72% of the study subjects had a GGN repeat length of 23.

Table 6 *Distribution of number of GGN repeats in the androgen receptor gene in prostate cancer cases and controls, China*

	Cases (n=187)		Controls (n=295)	
No. of GGN repeats	N	%	N	%
14	0	0.0	1	0.3
15	1	0.5	1	0.3
16	2	1.1	1	0.3
17	2	1.1	0	0.0
18	0	0.0	1	0.3
19	19	10.2	20	6.8
20	2	1.1	2	0.7
21	3	1.6	0	0.0
22	10	5.3	24	8.2
23	136	72.7	212	72.1
24	10	5.3	24	8.2
25	2	1.1	1	0.3
27	0	0.0	1	0.3
Median	23		23	

Risks of prostate cancer associated with CAG and GGN repeat lengths are shown in Table 7. When the number of CAG repeats was included in the model as a continuous variable, there was a 7% increase in the risk of prostate cancer for each decrement in length of one CAG repeat (OR=1.07, 95% CI 1.00-1.15). The risks associated with decrements of three and six repeats were 1.21 (95% 1.14-1.32) and

1.42 (95% CI 1.22-1.61), respectively. When the median repeat length was used to dichotomize study subjects, men with a CAG repeat length shorter than 23 had a 65% increased risk (OR=1.65, 95% CI 1.14-2.39), compared to men with a CAG repeat length of 23 or longer. Relative to the highest tertile of CAG repeat length (≥24), men in the second and first tertiles (22-23 and <22) had ORs of 1.45 and 1.55, respectively (P_{trend}=0.06).

Table 7 *Odds ratios (ORs)[a] and 95% confidence intervals (CIs) for prostate cancer in relation to the number of CAG and GGN repeats in the androgen receptor gene, China*

Number of CAG and GGN repeats	Cases No.	Controls No.	ORa	95% CI
No. of CAG repeats				
Continuous (per decrement of one CAG repeat)	190	300	1.07	1.00-1.15
Median				
≥23	74	154	1.00	-
<23	116	146	1.65	1.14-2.39
Tertile				
≥24	52	108	1.00	-
22-23	79	113	1.45	0.93-2.25
<22	59	79	1.55	0.96-2.49
			Linear trend	*p=0.06*
No. of GGN repeats				
Continuous (per decrement of one GGN repeat)	187	294	1.07	0.96-1.20
Median				
≥23	147	239	1.00	1.00
<23	39	56	1.12	0.71-1.78
Combined number of CAG and GGN repeats				
CAG≥23, GGN ≥23	53	120	1.00	-

CAG≥23, GGN < 23	19	29	1.48	0.76-2.88
CAG<23, GGN ≥23	94	115	1.85	1.21-2.82
CAG<23, GGN < 23	20	26	1.75	0.90-3.41

[a]Adjusted for age (continuous).

Similarly, men with a shorter GGN repeat length had a higher risk of prostate cancer. Each decrement of one GGN repeat length was associated with a 7% increase in risk (OR= 1.07, 95% CI 0.96-1.20). Men with a GGN repeat length shorter than the median length of 23 had a 12% increase in prostate cancer risk, compared to those with 23 or more repeats. Because more than 72% of the subjects had 23 GGN repeats, the ORs by tertiles for GGN repeats were not estimated.

Also shown in Table 7 are the ORs associated with combined categories of CAG and GGN repeat lengths. Men with both CAG and GGN repeat lengths shorter than 23 had a 75% elevated risk of prostate cancer. There was little correlation between the number of CAG and GGN repeats ($r=-0.03$, $p>0.05$).

The number of CAG or GGN repeats did not correlate with age, education, body mass index, waist-to-hip ratio, total calories, smoking, or drinking. These variables therefore were not included in the model for adjustment. The ORs were materially unchanged after further adjustment for benign prostatic hyperplasia (BPH), although the cases had a higher prevalence of BPH (57% vs. 23%) and there was a non-significant moderate association between CAG or GGN repeat lengths and BPH (data are reported separately). Associations of CAG or GGN repeat length were similar across all stages of disease at diagnosis (data not shown).

These results confirm that a shorter CAG repeat length is associated with an increased risk of clinically significant prostate cancer. A shorter length of GGN repeat also appears to increase the risk of prostate cancer, but the magnitude of excess risk was smaller.

The observed inverse association with AR polymorphisms is biologically

plausible, as laboratory studies have shown that a long polyglutamine chain (>30 repeats) in the AR gene is associated with androgen insensitivity and reduced AR transactivation activity (13,14AR). *In vitro* transfection studies also have demonstrated that elimination of the polyglutamine tracts results in elevated transcriptional activities (9,11,16AR). Clinical studies have suggested that alteration in the AR function, either through polymorphisms of CAG repeat length or somatic mutations, may be associated with tumor progression. For example, the progression from latent to clinically invasive prostate cancer is initially androgen-dependent, although some tumors later become androgen-independent (thus becoming non-responsive to hormonal treatment). Several non-germline-related changes of the AR gene, including amplification of the AR gene (usually a key step in the transition from a hormone-sensitive to a hormone-refractory state in prostate tumors) (31,32AR), AR somatic mutations (identified throughout transactivation, DNA binding, and ligand binding domains) (33,34AR), and contraction of CAG repeat length in cancer cells (32AR), have been shown to be associated with tumor aggressiveness, cancer progression, and failure of hormonal therapy. AR expression studies in the majority of prostate tumors, including those that have become refractory to hormonal therapy, also suggest that AR plays a key role in androgen-independent tumors (35,36AR).

The inverse relationship between CAG repeat length and AR transcriptional activity (thus androgen sensitivity) is the currently recognized underlying molecular mechanism by which AR polymorphisms modulate prostate cancer risk. Because transcriptional activation of the AR gene is influenced by not only polymorphisms in the AR gene but also a number of other factors, including tissue levels of dihydrotestosterone (DHT), estradiol, insulin-like growth factors, and AR coactivators (37-43AR), it is likely that these factors may also affect prostate cancer risk by mediating transcriptional activities. Several AR coactivators, including AR-associated proteins (ARA70, ARA55), AIB1 (Amplified in Breast Cancer -1), CBP (cyclic AMP responsive element binding protein), Rb, and BRCA1, have been shown to enhance AR- mediated transcriptional activity from 2- to 10-fold,

suggesting that *in vivo* coactivators are essential in attaining optimal AR transactivation in response to androgens (40-43AR).

It has been suggested that variations in CAG repeat length in the AR gene between populations may explain part of the large racial difference in prostate cancer risk and that a shorter CAG repeat length reported for African Americans may contribute to some of their higher risk of prostate cancer, although currently no data are available from this population. Our results confirm that, relative to western men, Chinese men do indeed have a longer CAG repeat length. For example, 22% of the 1,722 white men in two U.S. studies (17,18AR) had a CAG repeat length shorter than 20 vs. only 10% in our study and 55% reported for African Americans in a cross-sectional survey (26,27AR). Inverse associations have also been reported for Caucasians, suggesting that the underlying biological mechanism in various racial groups may be similar and that the polymorphisms of AR may be related, in part, to racial difference in prostate cancer risk.

The common polymorphism of the AR gene confers variable risk upon all individuals, which in turn may result in a much larger proportion of prostate cancer cases attributable to having fewer CAG repeats. Assuming that the CAG polymorphism association is causal, it is estimated that 25% (95% CI 9% to 41%) of the cases in Shanghai can be attributed to a CAG repeat length shorter than 23. Using the CAG repeat length distribution in the two U.S. studies among white men (17,18AR), it was estimated that 3-7%% of cases in the U.S. white men can be attributed to the CAG polymorphism (repeat length <23) and that this polymorphism alone potentially accounts for at least 5% of the difference in incidence between Chinese and U.S. men.

Similar to two previous studies (17,18AR), it was found that the number of GGN repeats clusters around 23 (in the study of Stanford et al., only the number of GGC repeats was counted and 15 was the peak of the repeat, which corresponds to 21 GGN repeats), and that a shorter GGN repeat length appears to be associated with a moderate increase in prostate cancer risk. Twenty-three GGN repeats may represent the coding sequence for optimal AR protein conformation and activity,

because over 70% of the study subjects in our study as well as in studies of western men had a GGN repeat length of 23.

Although it is well established that $(GGC)_n$ repeats in the polyglycine tract ($GGT_3GGG_1GGT_2GGC_n$) (SEQ ID NO:17) of the AR gene is polymorphic, to date there has been little information on variations in the GGG and GGT regions of the polyglycine tract, because these regions are GC-rich and technically it has been difficult to amplify these regions. Our study represents the first successful effort to sequence the exact codon usage and number of the GGN trinucleotide repeats in a large number of population-based samples. It was showed that GGT and GGG regions were quite stable and there were no variations in these two regions in all of the 481 DNA samples analyzed.

F. AIB1 references

Reference numbers throughout the application that have an "AIB1" associated with them refer to this list of references.

1. Hsing, A. W., Devesa, S. S., Jin, F., and Gao, Y. T. Rising incidence of prostate cancer in Shanghai, China. Cancer Epidemiol Biomarkers Prev, *7:* 83-84, 1998.

2. Hsing, A. W., Tsao, L., and Devesa, S. S. International trends and patterns of prostate cancer incidence and mortality. Int J Cancer, *85:* 60-67, 2000.

3. Hsing, A. W. and Devesa, S. S. Trends and patterns of prostate cancer: what do they suggest? Epidemiologic Reviews, 2001. *(in press)*

4. Ross, R. K., Pike, M. C., Coetzee, G. A., Reichardt, J. K., Yu, M. C., Feigelson, H., Stanczyk, F. Z., Kolonel, L. N., and Henderson, B. E. Androgen metabolism and prostate cancer: establishing a model of genetic susceptibility. Cancer Res, *58:* 4497-4504, 1998.

5. Ross, R. K., Bernstein, L., Lobo, R. A., Shimizu, H., Stanczyk, F. Z., Pike, M. C., and Henderson, B. E. 5-alpha-reductase activity and risk of prostate cancer among Japanese and US white and black males. Lancet, *339:* 887-889, 1992.

6. Shibata, A. and Whittemore, A. S. Genetic predisposition to prostate cancer: possible explanations for ethnic differences in risk. Prostate, *32:* 65-72, 1997.

7. Stanford, J. L., Just, J. J., Gibbs, M., Wicklund, K. G., Neal, C. L., Blumenstein, B. A., and Ostrander, E. A. Polymorphic repeats in the androgen receptor gene: molecular markers of prostate cancer risk. Cancer Res, *57:* 1194-1198, 1997.

8. Giovannucci, E., Stampfer, M. J., Krithivas, K., Brown, M., Dahl, D., Brufsky, A., Talcott, J., Hennekens, C. H., and Kantoff, P. W. The CAG repeat within the androgen receptor gene and its relationship to prostate cancer. Proc Natl Acad Sci U S A, *94:* 3320-3323, 1997.

9. Hsing, A. W., Gao, Y. T., Wu, G., Wang, X., Deng, J., Chen, Y. L., Sesterhenn, I. A., Mostofi, F. K., Benichou, J., and Chang, C. Polymorphic CAG and GGN repeat lengths in the androgen receptor gene and prostate cancer risk: a population-based case-control study in China. Cancer Res., *60:* 5111-5116, 2000.

10. Yeh, S., Miyamoto, H., Shima, H., and Chang, C. From estrogen to androgen receptor: a new pathway for sex hormones in prostate. Proc Natl Acad Sci U S A, *95:* 5527-5532, 1998.

11. Heinlein, C. A., Ting, H. J., Yeh, S., and Chang, C. Identification of ARA70 as a ligand-enhanced coactivator for the peroxisome proliferator-activated receptor gamma. J Biol Chem, *274:* 16147-16152, 1999.

12. Yeh, S., Kang, H. Y., Miyamoto, H., Nishimura, K., Chang, H. C., Ting, H. J., Rahman, M., Lin, H. K., Fujimoto, N., Hu, Y. C., Mizokami, A., Huang, K. E., and Chang, C. Differential induction of androgen receptor transactivation by different androgen receptor coactivators in human prostate cancer DU145 cells. Endocrine., *11:* 195-202, 1999.

13. Yeh, S. and Chang, C. Cloning and characterization of a specific coactivator, ARA70, for the androgen receptor in human prostate cells. Proc. Natl. Acad. Sci. U. S. A, *93:* 5517-5521, 1996.

14. Anzick, S. L., Kononen, J., Walker, R. L., Azorsa, D. O., Tanner, M. M., Guan, X. Y., Sauter, G., Kallioniemi, O. P., Trent, J. M., and Meltzer, P. S. AIB1, a steroid receptor coactivator amplified in breast and ovarian cancer. Science, *277:* 965-968, 1997.

15. Bautista, S., Valles, H., Walker, R. L., Anzick, S., Zeillinger, R., Meltzer, P., and Theillet, C. In breast cancer, amplification of the steroid receptor coactivator gene AIB1 is correlated with estrogen and progesterone receptor positivity. Clin Cancer Res, *4:* 2925-2929, 1998.

16. Hsing, A. W., Chen, C., Chokkalingam, A. P., Gao, Y. T., Dightman, D. A., Nguyen, H. T., Deng, J., Cheng, J., Sesterhenn, I. A., Mostofi, F. K., Stanczyk, F. Z., and Reichardt, J. K. V. Polymorphic markers in the SRD5A2 gene and prostate cancer risk: a population-based case-control study. Cancer Epidemiol Biomarkers Prev, 2001. *(in press)*

17. Hsing, A. W., Deng, J., Sesterhenn, I. A., Mostofi, F. K., Stanczyk, F. Z., Benichou, J., Xie, T., and Gao, Y. T. Body size and prostate cancer: a population-based case-control study in China. Cancer Epidemiol. Biomarkers Prev., *9:* 1335-1341, 2000.

18. Hsing, A. W., Chua, S., Jr., Gao, Y. T., Gentzschein, E., Chang, L., Deng, J., and Stanczyk, F. Z. Prostate cancer risk and serum levels of insulin and leptin: a population-based study. J. Natl. Cancer Inst., *93:* 783-789, 2001.

19. Shirazi, S. K., Bober, M. A., and Coetzee, G. A. Polymorphic exonic CAG microsatellites in the gene amplified in breast cancer (AIB1 gene). Clin Genet, *54:* 102-103, 1998.

20. Platz, E. A., Giovannucci, E., Brown, M., Cieluch, C., Shepard, T. F., Stampfer, M. J., and Kantoff, P. W. Amplified in breast cancer-1 glutamine repeat and prostate cancer risk. Prostate J., *2:* 27-32, 2000.

21. Rebbeck, T. R., Wang, Y., Kantoff, P. W., Krithivas, K., Neuhausen,

S. L., Godwin, A. K., Daly, M. B., Narod, S. A., Brunet, J. S., Vesprini, D., Garber, J. E., Lynch, H. T., Weber, B. L., and Brown, M. Modification of brca1- and brca2-associated breast cancer risk by aib1 genotype and reproductive history. Cancer Res., *61:* 5420-5424, 2001.

22. Chang, C. S., Kokontis, J., and Liao, S. T. Structural analysis of complementary DNA and amino acid sequences of human and rat androgen receptors. Proc. Natl. Acad. Sci. U. S. A, *85:* 7211-7215, 1988.

23. Breslow, N. E. and Day, N. E. Statistical methods in cancer research. Volume I - The analysis of case-control studies. IARC Sci Publ, *32:* 5-338, 1980.

24. Thenot, S., Charpin, M., Bonnet, S., and Cavailles, V. Estrogen receptor cofactors expression in breast and endometrial human cancer cells. Mol. Cell Endocrinol., *156:* 85-93, 1999.

25. Bubendorf, L,, Kononen, J., Koivisto, P., Schraml, P., Moch, H., Gasser, T. C., Willi, N., Mihatsch, M. J., Sauter, G., and Kallioniemi, O. P. Survey of gene amplifications during prostate cancer progression by high-throughout fluorescence in situ hybridization on tissue microarrays. Cancer Res., *59:* 803-806, 1999.

26. Hayashi, Y., Yamamoto, M., Ohmori, S., Kikumori, T., Imai, T., Funahashi, H., and Seo, H. Polymorphism of homopolymeric glutamines in coactivators for nuclear hormone receptors. Endocr J, *46:* 279-284, 1999.

27. McKenna, N. J., Xu, J., Nawaz, Z., Tsai, S. Y., Tsai, M. J., and O'Malley, B. W. Nuclear receptor coactivators: multiple enzymes, multiple complexes, multiple functions. J. Steroid Biochem. Mol. Biol., *69:* 3-12, 1999.

28. Heery, D. M., Kalkhoven, E., Hoare, S., and Parker, M. G. A signature motif in transcriptional co-activators mediates binding to nuclear receptors. Nature, *387:* 733-736, 1997.

29. Smith, J. R., Freije, D., Carpten, J. D., Gronberg, H., Xu, J., Isaacs, S. D., Brownstein, M. J., Bova, G. S., Guo, H., Bujnovszky, P., Nusskern, D. R., Damber, J. E., Bergh, A., Emanuelsson, M., Kallioniemi, O. P., Walker-Daniels, J., Bailey-Wilson, J. E., Beaty, T. H., Meyers, D. A., Walsh, P. C., Collins, F. S., Trent, J. M., and Isaacs, W. B. Major susceptibility locus for prostate cancer on chromosome 1 suggested by a genome-wide search. Science, *274:* 1371-1374, 1996.

G. AR References

Reference numbers throughout the application that have an "AR" associated with them refer to this list of references.

1. Parkin, D.M., Whelan, S.L., Ferlay, J., Raymond, L., and Young, L. (eds). Cancer incidence in Five Continents. Vol II. IARC Scientific Publications, 143, International Agency for Research on Cancer, Lyon, 1997.

2. Hsing, A.W., Tsao, L., and Devesa, S.S. International trends and patterns of prostate cancer incidence and mortality. Int. J. Cancer, 85: 60-67, 2000.

3. Hsing, A.W., Devesa, S.S., Jin, F., and Gao, Y.-T. Rising incidence of prostate cancer in Shanghai, China. Cancer Epidemiol. Biomarkers Prev., 7: 83-84, 1998.

4. Shimizu, H., Ross, R.K., and Bernstein, L. Possible underestimation of the incidence rate of prostate cancer in Japan. Jpn. J. Cancer Res., 82: 483-485, 1991.

5. Breslow, N., Chan, C.W., Dhom, G., Drury, R.A., Franks, L.M., Gellei, B., Lee, Y.S., Lundberg, S., Sparke, B., Sternby, N.H., and Tulinius, H. Latent carcinoma of the prostate of autopsy in seven areas. Int. J. Cancer, 20: 680-688, 1977.

6. Roy, A.K., Lavrovsky, Y., Song, C.S., Chen, S., Jung, M.H., Velu, N.K., Bi, B.Y., and Chatterjee, B. Regulation of androgen action. Vitam. Horm., 55: 309-352, 1999.

7. Trapman, J., and Brinkmann, A.O. The androgen receptor in prostate cancer. Path. Res. Pract., 192: 752-760, 1996.

8. McPhaul, M.J. Molecular defects of the androgen receptor. J. Steroid Biochem. Mol. Biol., 69: 315-322,1999.

9. Chamberlain, N.L., Driver, E.D., and Miesfeld, R.L.T. The length and location of CAG trinucleotide repeats in the androgen receptor N-terminal domain affect transactivation function. Nucleic Acids Res., 22: 3181-3186, 1994.

10. Choong, C.S., Kemppainen, J.A., Zhou, Z.X., and Wilson, E.M. Reduced androgen receptor gene expression with first exon CAG repeat expansion. Mol. Endocrinol., 10: 1527-1535, 1996.

11. Tut, T.G., Ghadessy, F.J., Trifiro, M.A., Pinsky, L., and Yong, E.L. Long polyglutamine tracts in the androgen receptor are associated with reduced trans-activation, impaired sperm production, and male infertility. J. Clin. Endocrinol. Metab., 82: 3777-3782, 1997.

12. Zirkin B.R. Spermatogenesis: its regulation by testosterone and FSH. Semin. Cell Dev. Biol., 9: 417-421, 1998.

13. Fischbeck, K.H., Lieberman, A., Bailey, C.K., Abel, A., and Merry D.E. Androgen receptor mutation in Kennedy's disease. Philos. Trans. R. R. Soc. Lond. B. Biol. Sci., 354: 1075-1078, 1999.

14. Brooks, B.P., and Fischbeck, K.H. Spinal and bulbar muscular atrophy: a trinucleotide-repeat expansion neurodegenerative disease. Trends Neurosci., 18: 459-461, 1995.

15. Kazemi-Esfarjani, P., Trifiro, M.A., and Pinsky, L. Evidence for a repressive function of the long polyglutamine tract in the human androgen receptor: possible pathogenetic relevance for the (CAG)n-expanded neuronopathies. Hum. Mol. Genet., 4: 523-527, 1995.

16. Gao, T., Marcelli, M., and McPhaul, M.J. Transcriptional activation and transient expression of the human androgen receptor. J. Steroid Biochem. Mol. Biol., 59:9-20, 1996.

17. Stanford, J.L., Just, J.J., Gibbs, M., Wicklund, K.G., Neal, C.L., Blumenstein, B.A., and Ostrander, E.A. Polymorphic repeats in the androgen receptor gene: molecular markers of prostate cancer risk. Cancer Res., 57: 1194-1198, 1997.

18. Giovannucci, E., Stampfer, M.J., Krithivas, K., Brown, M., Dahl, D., Brufsky, A., Talcott, J., Hennekens, C.H., and Kantoff, P.W. The CAG repeat within the androgen receptor gene and its relationship to prostate cancer. Proc. Natl. Acad. Sci., 94: 3320-3323, 1997.

19. Ingles, S.A., Ross, R.K., Yu, M.C., Irvine, R.A., La Pera, G., Haile, R.W.,

and Coetzee, G.A Association of prostate cancer risk with genetic polymorphisms in vitamin D receptor and androgen receptor. J. Natl. Cancer Inst., 89: 166-170, 1997.

20. Platz, E.A., Giovannucci, E., Dahl, D.M., Krithivas, K., Hennekens, C.H., Brown, M., Stampfer, M.J., and Kantoff, P.W. The androgen receptor gene GGN microsatellite and prostate cancer risk. Cancer Epidemiol. Biomarkers Prev., 7: 379-384, 1998.

21. Kantoff, P., Giovannucci, E., and Brown, M. The androgen receptor CAG repeat polymorphism and its relationship to prostate cancer. Biochim. Biophys. Acta., 1378: C1-C5, 1998.

22. Hakimi, J.M., Schoenberg, M.P., Rondinelli, R.H., Piantadosi, S., and Barrack, E.R. Androgen receptor variants with short glutamine or glycine repeats may identify unique subpopulations of men with prostate cancer. Clin. Cancer Res., 3: 1599-1608, 1997.

23. Sartor, O., Zheng, Q., and Eastham, J.A. Androgen receptor gene CAG repeat length varies in a race-specific fashion in men without prostate cancer. Urology, 53: 378-380, 1999.

24. Hardy, D.O., Scher, H.I., Bogenreider, T., Sabbatini, P., Zhang, Z.F., Nanus, D.M., and Catterall, J.F. Androgen receptor CAG repeat lengths in prostate cancer: correlation with age of onset. Clin. Endocrinol. Metab., 81: 4400-4405, 1996.

25. Bratt, O., Borg, A., Kristoffersson, U., Lundgren, R., Zhang, Q.X., Olsson, H. CAG repeat length in the androgen receptor gene is related to age at prostate cancer diagnosis and response to endocrine therapy, but not to risk of prostate cancer. Br. J. Cancer, 81: 672-676, 1999.

26. Irvine, R.A., Yu, M.C., Ross, R.K., and Coetzee, G.A. The CAG and GGC microsatellites of the androgen receptor gene are in linkage disequilibrium in men with prostate cancer. Cancer Res., 55: 1937-1940, 1995.

27. Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics;12: 241-253, 1992.

28. Hsing, A.W., Deng, J., Tong, X., Sesterhenn, I., Mostofi, F.K., and Gao, Y-T. Body size and prostate cancer: a population-based case-control study in Shanghai, China (Submitted).

29. Breslow, N.E. and Day, N.E. Statistical methods in cancer research. Vol I. The analysis of case-control studies. IARC Scientific Publication 32. IARC. Lyon,1980.

30. Correa-Cerro, L., Wohr, G., Haussler, J., Berthon, P., Drelon, E., Mangin, P., Fournier, G., Cussenot, O., Kraus, P., Just, W., Paiss, T., Cantu, J.M., Vogel, W. (CAG)nCAA and GGN repeats in the human androgen receptor gene are not associated with prostate cancer in a French-German population. Eur. J. Hum. Genet., 7: 357-362, 1999.

31. Wallen MJ, Linja M, Kaartinen K, Schleutker J, Visakorpi T. Androgen receptor gene mutations in hormone-refractory prostate cancer. J. Pathol., 189: 559-563, 1999.

32. Koivisto, P.A., Rantala, I. Amplification of the androgen receptor gene is associated with P53 mutation in hormone-refractory recurrent prostate cancer. J.

Pathol.,187: 237-241, 1999.

33. Culig, Z., Hobisch, A., Cronauer, M.V., Cato, A.C., Hittmair, A., Radmayr, C., Eberle, J., Bartsch, G., and Klocker, H. Mutant androgen receptor detected in an advanced-stage prostatic carcinoma is activated by adrenal androgens and progesterone. Mol. Endocrinol., 7: 1541-1550, 1997.

34. Schoenberg, M.P., Hakimi, J.M., Wang, S., Bova, G.S., Epstein, J.I., Fischbeck, K.H., Isaacs, W.B., Walsh, P.C., Barrack, E.R. Microsatellite mutation (CAG24-->18) in the androgen receptor gene in human prostate cancer. Biochem. Biophys. Res. Commun.,198: 74-80, 1994.

35. Jenster, G. The role of the androgen receptor in the development and progression of prostate cancer. Semin. Oncol., 26: 407-21, 1999.

36. Koivisto, P., Kononen, J., Palmberg, C., Tammela, T., Hyytinen, E., Isola, J., Trapman, J., Cleutjens, K., Noordzij, A., Visakorpi, T., Kallioniemi, O.P. Androgen receptor gene amplification: a possible molecular mechanism for androgen deprivation therapy failure in prostate cancer. Cancer Res., 57: 314-319, 1997.

37. Culig, Z., Hobisch, A., Cronauer, M.V., Hittmair, A., Radmayr, C., Bartsch, G., and Klocker, H. Activation of the androgen receptor by polypeptide growth factors and cellular regulators. World J. Urol., 13: 285-289, 1995.

38. McAbee, M.D., and Doncarlos, L.L. Estrogen, but not androgens, regulates androgen receptor messenger ribonucleic acid expression in the developing male rat forebrain. Endocrinology, 140: 3674-3681, 1999.

39. Gupta, C. Modulation of androgen receptor (AR)-mediated transcriptional activity by EGF in the developing mouse reproductive tract primary cells. Mol. Cell. Endocrinol., 152: 169-178, 1999.

40. Cude, K.J., Dixon, S.C., Guo, Y., Lisella, J., Figg, W.D. The androgen receptor: genetic considerations in the development and treatment of prostate cancer. J. Mol. Med., 77: 419-426, 1999.

41. Yeh, S., and Chang, C. Cloning and characterization of a specific coactivator, ARA70, for the androgen receptor in human prostate cells. Proc. Natl. Acad. Sci., 93: 5517-5521, 1997.

42. Yeh, S., Miyamoto, H., Shima, H., and Chang, C. From estrogen to androgen receptor: a new pathway for sex hormones in prostate. Proc. Natl. Acad. Sci., 95: 5527-5532, 1998.

43. Anzick, S.L., Kononen, J., Walker, R.L., Azorsa, D.O., Tanner, M.M., Guan, X.Y., Sauter, G., Kallioniemi, O.P., Trent, J.M., and Meltzer, P.S. AIB1, a steroid receptor coactivator amplified in breast and ovarian cancer. Science, 277: 965-968, 1997.

H. Sequences

SEQ ID NO:1 first primer
5'-TCA TCA CCT CCG ACA ACA GAG G-3'

SEQ ID NO:2 second primer
5'-TAT GGA AAC TGT TGC GGA GGA G-3'

SEQ ID NO:3 complement to first primer
5'-CCT CTG TTG TCG GAG GTG ATG A-3'

SEQ ID NO:4 complement to second primer
5'-CTC CTC CGC AAC AGT TTC CAT A -3'

SEQ ID NO:5 general CAG/CAA sequence
5'-$(CAG)_N$ CAA $(CAG)_N$ $(CAACAG)_4$ CAG CAA CAG CAG CAA-3'
Nucleotides 1-3 and 6-9 as a triplets must be at least one triple but can be any multiple of triplets.

SEQ ID NO:6 one particular 29 mer sequence
5'-CAG CAG CAG CAG CAG CAG CAA CAG CAG CAG CAG CAG CAG CAG CAG CAG CAA CAG CAA CAG CAA CAG CAA CAG CAG CAA CAG CAG CAA-3'

SEQ ID NO:7 first androgen receptor CAG primer
5'-GCT CTG GGA CGC AACCTCTCT-3'

SEQ ID NO:8 second androgen receptor CAG primer
5'-GCA GCG ACT ACC GCA TCA TCA-3'

SEQ ID NO:9 complement to first androgen receptor CAG primer
5'-AGAGAGGTTGCGTCCCAGAGC-3'

SEQ ID NO:10 complement to second androgen receptor CAG primer
5'-TGATGATGCGGTAGTCGCTGC-3'

SEQ ID NO:11 first androgen receptor GGN primer
5'ACCCTCAGCCGCCGCTTCCTCATC-3''

SEQ ID NO:12 second androgen receptor GGN primer
5'-CTGGGATAGGGCACTCTGCTCAAC-3'

SEQ ID NO:13 complement to first androgen receptor GGN primer
5'-GATGAGGAAGCGGCGGCTGAGGGT-3'

SEQ ID NO:14 complement to second androgen receptor GGN primer
5'-GTTGAGCAGAGTGCCCTATCCCAG-3'

SEQ ID NO:15
5'-CGGG-GTAAGGGAAGTAGGTGGAAG-3'

SEQ ID NO:16
and 5'-CTCTACGATGGGCTTGGGGAGAAC-3'

SEQ ID NO:17
$GGT_3GGG_1GGT_2GGC_n$

SEQ ID NO:18 AIB1 sequence Genbank accession number XM_030032

1 cggcagcggc tgcggcttag tcggtggcgg ccggcggcgg ctgcgggctg agcggcgagt
61 ttccgattta aagctgagct gcgaggaaaa tggcggcggg aggatcaaaa tacttgctgg
121 atggtggact cagagaccaa taaaaataaa ctgcttgaac atcctttgac tggttagcca
181 gttgctgatg tatattcaag atgagtggat taggagaaaa cttggatcca ctggccagtg
241 attcacgaaa acgcaaattg ccatgtgata ctccaggaca aggtcttacc tgcagtggtg
301 aaaaacggag acgggagcag gaaagtaaat atattgaaga attggctgag ctgatatctg
361 ccaatcttag tgatattgac aatttcaatg tcaaaccaga taaatgtgcg attttaaagg
421 aaacagtaag acagatacgt caaataaaag agcaaggaaa aactatttcc aatgatgatg
481 atgttcaaaa agccgatgta tcttctacag ggcagggagt tattgataaa gactccttag
541 gaccgctttt acttcaggca ttggatggtt tcctatttgt ggtgaatcga gacggaaaca
601 ttgtatttgt atcagaaaat gtcacacaat acctgcaata taagcaagag gacctggtta
661 acacaagtgt ttacaatatc ttacatgaag aagacagaaa ggattttctt aagaatttac
721 caaaatctac agttaatgga gtttcctgga caaatgagac ccaaagacaa aaaagccata
781 catttaattg ccgtatgttg atgaaaacac cacatgatat tctggaagac ataaacgcca
841 gtcctgaaat gcgccagaga tatgaaacaa tgcagtgctt tgccctgtct cagccacgag
901 ctatgatgga ggaaggggaa gatttgcaat cttgtatgat ctgtgtggca cgccgcatta
961 ctacaggaga aagaacattt ccatcaaacc ctgagagctt tattaccaga catgatcttt
1021 caggaaaggt tgtcaatata gatacaaatt cactgagatc ctccatgagg cctggctttg
1081 aagatataat ccgaaggtgt attcagagat tttttagtct aaatgatggg cagtcatggt
1141 cccagaaacg tcactatcaa gaagcttatc ttaatggcca tgcagaaacc ccagtatatc
1201 gattctcgtt ggctgatgga actatagtga ctgcacagac aaaaagcaaa ctcttccgaa
1261 atcctgtaac aaatgatcga catggctttg tctcaaccca cttccttcag agagaacaga
1321 atggatatag accaaaccca aatcctgttg gacaagggat tagaccacct atggctggat
1381 gcaacagttc ggtaggcggc atgagtatgt cgccaaacca aggcttacag atgccgagca
1441 gcagggccta tggcttggca gaccctagca ccacagggca gatgagtgga gctaggtatg
1501 ggggttccag taacatagct tcattgaccc ctgggccagg catgcaatca ccatcttcct
1561 accagaacaa caactatggg ctcaacatga gtagcccccc acatgggagt cctggtcttg
1621 ccccaaacca gcagaatatc atgatttctc ctcgtaatcg tgggagtcca aagatagcct
1681 cacatcagtt ttctcctgtt gcaggtgtgc actctcccat ggcatcttct ggcaatactg
1741 ggaaccacag cttttccagc agctctctca gtgccctgca agccatcagt gaaggtgtgg
1801 ggacttccct tttatctact ctgtcatcac caggccccaa attggataac tctcccaata
1861 tgaatattac ccaaccaagt aaagtaagca atcaggattc caagagtcct ctgggctttt
1921 attgcgacca aaatccagtg gagagttcaa tgtgtcagtc aaatagcaga gatcacctca
1981 gtgacaaaga aagtaaggag agcagtgttg agggggcaga gaatcaaagg ggtcctttgg
2041 aaagcaaagg tcataaaaaa ttactgcagt tacttacctg ttcttctgat gaccggggtc
2101 attcctcctt gaccaactcc cccctagatt caagttgtaa agaatcttct gttagtgtca
2161 ccagcccctc tggagtctcc tcctctacat ctggaggagt atcctctaca tccaatatgc
2221 atgggtcact gttacaagag aagcaccgga ttttgcacaa gttgctgcag aatgggaatt
2281 caccagctga ggtagccaag attactgcag aagccactgg gaaagacacc agcagtataa
2341 cttcttgtgg ggacggaaat gttgtcaagc aggagcagct aagtcctaag aagaaggaga

2401 ataatgcact tcttagatac ctgctggaca gggatgatcc tagtgatgca ctctctaaag
2461 aactacagcc ccaagtggaa ggagtggata ataaaatgag tcagtgcacc agctccacca
2521 ttcctagctc aagtcaagag aaagacccta aaattaagac agagacaagt gaagagggat
2581 ctggagactt ggataatcta gatgctattc ttggtgatct gactagttct gacttttaca
2641 ataattccat atcctcaaat ggtagtcatc tggggactaa gcaacaggtg tttcaaggaa
2701 ctaattctct gggtttgaaa agttcacagt ctgtgcagtc tattcgtcct ccatataacc
2761 gagcagtgtc tctggatagc cctgtttctg ttggctcaag tcctccagta aaaaatatca
2821 gtgctttccc catgttacca aagcaaccca tgttgggtgg gaatccaaga atgatggata
2881 gtcaggaaaa ttatggctca agtatgggtg ggccaaaccg aaatgtgact gtgactcaga
2941 ctccttcctc aggagactgg ggcttaccaa actcaaaggc cggcagaatg gaacctatga
3001 attcaaactc catgggaaga ccaggaggag attataatac ttctttaccc agacctgcac
3061 tgggtggctc tattcccaca ttgcctcttc ggtctaatag cataccaggt gcgagaccag
3121 tattgcaaca gcagcagcag atgcttcaaa tgaggcctgg tgaaatcccc atgggaatgg
3181 gggctaatcc ctatggccaa gcagcagcat ctaaccaact gggttcctgg cccgatggca
3241 tgttgtccat ggaacaagtt tctcatggca ctcaaaatag gcctcttctt aggaattccc
3301 tggatgatct tgttgggcca ccttccaacc tggaaggcca gagtgacgaa agagcattat
3361 tggaccagct gcacactctt ctcagcaaca cagatgccac aggcctggaa gaaattgaca
3421 gagctttggg cattcctgaa cttgtcaatc agggacaggc attagagccc aaacaggatg
3481 ctttccaagg ccaagaagca gcagtaatga tggatcagaa ggcaggatta tatggacaga
3541 catacccagc acaggggcct ccaatgcaag gaggctttca tcttcaggga caatcaccat
3601 cttttaactc tatgatgaat cagatgaacc agcaaggcaa ttttcctctc caaggaatgc
3661 acccacgagc caacatcatg agaccccgga caaacacccc caagcaactt agaatgcagc
3721 ttcagcagag gctgcagggc cagcagtttt tgaatcagag ccgacaggca cttgaattga
3781 aaatggaaaa ccctactgct ggtggtgctg cggtgatgag gcctatgatg cagccccagc
3841 agggttttct taatgctcaa atggtcgccc aacgcagcag agagctgcta agtcatcact
3901 tccgacaaca gagggtggct atgatgatgc agcagcagca gcagcagcaa cagcagcagc
3961 agcagcagca gcagcagcaa cagcaacagc aacagcaaca gcagcaacag cagcaaaccc
4021 aggccttcag cccacctcct aatgtgactg cttcccccag catggatggg cttttggcag
4081 gacccacaat gccacaagct cctccgcaac agtttccata tcaaccaaat tatggaatgg
4141 gacaacaacc agatccagcc tttggtcgag tgtctagtcc tcccaatgca atgatgtcgt
4201 caagaatggg tccctcccag aatcccatga tgcaacaccc gcaggctgca tccatctatc
4261 agtcctcaga aatgaagggc tggccatcag gaaatttggc caggaacagc tccttttccc
4321 agcagcagtt tgcccaccag gggaatcctg cagtgtatag tatggtgcac atgaatggca
4381 gcagtggtca catgggacag atgaacatga accccatgcc catgtctggc atgcctatgg
4441 gtcctgatca gaatactgct gacatctctg caccaggacc tcttaaggaa accactgtac
4501 aaatgacact gcactaggat tattgggaag gaatcattgt tccaggcatc catcttggaa
4561 gaaaggacca gctttgagct ccatcaaggg tattttaagt gatgtcattt gagcaggact
4621 ggatttaag ccgaagggca atatctacgt gtttttcccc cctccttctg ctgtgtatca
4681 tggtgttcaa aacagaaatg ttttttggca ttccacctcc tagggatata attctggaga
4741 catggagtgt tactgatcat aaaacttttg tgtcactttt ttctgccttg ctagccaaaa
4801 tctcttaaat acacgtaggt gggccagaga acattggaag aatcaagaga gattagaata
4861 tctggtttct ctagttgcag tattggacaa agagcatagt cccagccttc aggtgtagta
4921 gttctgtgtt gacccttttgt ccagtggaat tggtgattct gaattgtcct ttactaatgg
4981 tgttgagttg ctctgtccct attatttgcc ctaggctttc tcctaatgaa ggttttcatt
5041 tgccattcat gtcctgtaat acttcacctc caggaactgt catggatgtc caaatggctt

5101 tgcagaaagg aaatgagatg acagtattta atcgcagcag tagcaaactt ttcacatgct
5161 aatgtgcagc tgagtgcact ttatttaaaa agaatggata aatgcaatat tcttgaggtc
5221 ttgagggaat agtgaaacac attcctggtt tttgcctaca cttacgtgtt agacaagaac
5281 tatgattttt ttttttaaag tactggtgtc acccttgcc tatatggtag agcaataatg
5341 ctttttaaaa ataaacttct gaaaacccaa ggccaggtac tgcattctga atcagaatct
5401 cgcagtgttt ctgtgaatag attttttgt aaatatgacc tttaagatat tgtattatgt
5461 aaaatatgta tataccttt tttgtaggtc acaacaactc atttttacag agtttgtgaa
5521 gctaaatatt taacattgtt gatttcagta agctgtgtgg tgaggctacc agtggaagag
5581 acatcccttg acttttgtgg cctgggggag gggtagtgct ccacagcttt tccttcccca
5641 ccccccagcc ttagatgcct cgctcttttc aatctcttaa tctaaatgct ttttaaagag
5701 attatttgtt tagatgtagg cattttaatt ttttaaaaat tcctctacca gaactaagca
5761 ctttgttaat ttgggggggaa agaatagata tggggaaata aacttaaaaa aaaatcagga
5821 atttaaaaaa acgagcaatt tgaagagaat cttttggatt ttaagcagtc cgaaataata
5881 gcaattcatg ggctgtgtgt gtgtgtgtat gtgtgtgtgt gtgtgtgtat gtttaattat
5941 gttacctttt catccccttt aggagcgttt tcagattttg gttgctaaga cctgaatccc
6001 atattgagat ctcgagtaga atccttggtg tggtttctgg tgtctgctca gctgtcccct
6061 cattctacta atgtgatgct ttcattatgt ccctgtggat tagaatagtg tcagttattt
6121 cttaagtaac tcagtaccca gaacagccag ttttactgtg attcagagcc acagtctaac
6181 tgagcacctt ttaaacccct ccctcttctg ccccctacca cttttctgct gttgcctctc
6241 tttgacacct gttttagtca gttgggagga agggaaaaat caagtttaat tccctttatc
6301 tgggttaatt catttggttc aaatagttga cggaattggg tttctgaatg tctgtgaatt
6361 tcagaggtct ctgctagcct tggtatcatt ttctagcaat aactgagagc cagttaattt
6421 taagaatttc acacatttag ccaatctttc tagatgtctc tgaaggtaag atcatttaat
6481 atctttgata tgcttacgag taagtgaatc ctgattattt ccagacccac caccagagtg
6541 gatcttattt tcaaagcagt atagacaatt atgagtttgc cctctttccc ctaccaagtt
6601 caaaatatat ctaagaaaga ttgtaaatcc gaaaacttcc attgtagtgg cctgtgcttt
6661 tcagatagta tactctcctg tttggagaca gaggaagaac caggtcagtc tgtctctttt
6721 tcagctcaat tgtatctgac ccttctttaa gttatgtgtg tggggagaaa tagaatggtg
6781 ctcttatctt tcttgacttt aaaaaaatta ttaaaaacaa aaaaaaaata aa

SEQ ID NO:19

MSGLGENLDPLASDSRKRKLPCDTPGQGLTCSGEKRRREQESKY
IEELAELISANLSDIDNFNVKPDKCAILKETVRQIRQIKEQGKTISNDDDVQKADVSS
TGQGVIDKDSLGPLLLQALDGFLFVVNRDGNIVFVSENVTQYLQYKQEDLVNTSVYNI
LHEEDRKDFLKNLPKSTVNGVSWTNETQRQKSHTFNCRMLMKTPHDILEDINASPEMR
QRYETMQCFALSQPRAMMEEGEDLQSCMICVARRITTGERTFPSNPESFITRHDLSGK
VVNIDTNSLRSSMRPGFEDIIRRCIQRFFSLNDGQSWSQKRHYQEAYLNGHAETPVYR
FSLADGTIVTAQTKSKLFRNPVTNDRHGFVSTHFLQREQNGYRPNPNPVGQGIRPPMA
GCNSSVGGMSMSPNQGLQMPSSRAYGLADPSTTGQMSGARYGGSSNIASLTPGPGMQS
PSSYQNNNYGLNMSSPPHGSPGLAPNQQNIMISPRNRGSPKIASHQFSPVAGVHSPMA
SSGNTGNHSFSSSSLSALQAISEGVGTSLLSTLSSPGPKLDNSPNMNITQPSKVSNQD
SKSPLGFYCDQNPVESSMCQSNSRDHLSDKESKESSVEGAENQRGPLESKGHKKLLQL
LTCSSDDRGHSSLTNSPLDSSCKESSVSVTSPSGVSSSTSGGVSSTSNMHGSLLQEKH
RILHKLLQNGNSPAEVAKITAEATGKDTSSITSCGDGNVVKQEQLSPKKKENNALLRY
LLDRDDPSDALSKELQPQVEGVDNKMSQCTSSTIPSSSQEKDPKIKTETSEEGSGDLD
NLDAILGDLTSSDFYNNSISSNGSHLGTKQQVFQGTNSLGLKSSQSVQSIRPPYNRAV
SLDSPVSVGSSPPVKNISAFPMLPKQPMLGGNPRMMDSQENYGSSMGGPNRNVTVTQT
PSSGDWGLPNSKAGRMEPMNSNSMGRPGGDYNTSLPRPALGGSIPTLPLRSNSIPGAR
PVLQQQQQMLQMRPGEIPMGMGANPYGQAAASNQLGSWPDGMLSMEQVSHGTQNRPLL

RNSLDDLVGPPSNLEGQSDERALLDQLHTLLSNTDATGLEEIDRALGIPELVNQGQAL
EPKQDAFQGQEAAVMMDQKAGLYGQTYPAQGPPMQGGFHLQGQSPSFNSMMNQMNQQG
NFPLQGMHPRANIMRPRTNTPKQLRMQLQQRLQGQQFLNQSRQALELKMENPTAGGAA
VMRPMMQPQQGFLNAQMVAQRSRELLSHHFRQQRVAMMMQQQQQQQQQQQQQQQQQQ
Q
QQQQQQQQQQTQAFSPPPNVTASPSMDGLLAGPTMPQAPPQQFPYQPNYGMGQQPDPA
FGRVSSPPNAMMSSRMGPSQNPMMQHPQAASIYQSSEMKGWPSGNLARNSSFSQQQFA
HQGNPAVYSMVHMNGSSGHMGQMNMNPMPMSGMPMGPDQNTADISAPGPLKETTVQMT
LH

SEQ ID NO:20

1: NM_000044

1 cgagatcccg gggagccagc ttgctgggag agcgggacgg tccggagcaa gcccacaggc
61 agaggaggcg acagagggaa aaagggccga gctagccgct ccagtgctgt acaggagccg
121 aagggacgca ccacgccagc cccagcccgg ctccagcgac agccaacgcc tcttgcagcg
181 cggcggcttc gaagccgccg cccggagctg ccctttcctc ttcggtgaag tttttaaaag
241 ctgctaaaga ctcggaggaa gcaaggaaag tgcctggtag gactgacggc tgcctttgtc
301 ctcctcctct ccacccccgcc tccccccacc ctgccttccc cccctccccc gtcttctctc
361 ccgcagctgc ctcagtcggc tactctcagc caacccccct caccaccctt ctccccaccc
421 gcccccccgc ccccgtcggc ccagcgctgc cagcccgagt ttgcagagag gtaactccct
481 ttggctgcga gcgggcgagc tagctgcaca ttgcaaagaa ggctcttagg agccaggcga
541 ctggggagcg gcttcagcac tgcagccacg acccgcctgg ttagaattcc ggcggagaga
601 accctctgtt ttcccccact ctctctccac ctcctcctgc cttccccacc ccgagtgcgg
661 agcagagatc aaaagatgaa aaggcagtca ggtcttcagt agccaaaaaa caaaacaaac
721 aaaaacaaaa aagccgaaat aaaagaaaaa gataataact cagttcttat ttgcacctac
781 ttcagtggac actgaatttg gaaggtggag gattttgttt ttttctttta agatctgggc
841 atcttttgaa tctacccttc aagtattaag agacagactg tgagcctagc agggcagatc
901 ttgtccaccg tgtgtcttct tctgcacgag actttgaggc tgtcagagcg ctttttgcgt
961 ggttgctccc gcaagtttcc ttctctggag cttcccgcag gtgggcagct agctgcagcg
1021 actaccgcat catcacagcc tgttgaactc ttctgagcaa gagaagggga ggcggggtaa
1081 gggaagtagg tggaagattc agccaagctc aaggatggaa gtgcagttag ggctgggaag
1141 ggtctaccct cggccgccgt ccaagaccta ccgaggagct ttccagaatc tgttccagag
1201 cgtgcgcgaa gtgatccaga acccgggccc caggcaccca gaggccgcga gcgcagcacc
1261 tcccggcgcc agtttgctgc tgctgcagca gcagcagcag cagcagcagc agcagcagca
1321 gcagcagcag cagcagcagc agcagcaaga gactagcccc aggcagcagc agcagcagca
1381 gggtgaggat ggttctcccc aagcccatcg tagaggcccc acaggctacc tggtcctgga
1441 tgaggaacag caaccttcac agccgcagtc ggccctggag tgccaccccg agagaggttg
1501 cgtcccagag cctggagccg ccgtggccgc cagcaagggg ctgccgcagc agctgccagc
1561 acctccggac gaggatgact cagctgcccc atccacgttg tccctgctgg gccccacttt
1621 ccccggctta agcagctgct ccgctgacct taaagacatc ctgagcgagg ccagcaccat
1681 gcaactcctt cagcaacagc agcaggaagc agtatccgaa ggcagcagca gcgggagagc
1741 gagggaggcc tcgggggctc ccacttcctc caaggacaat tacttagggg gcacttcgac
1801 catttctgac aacgccaagg agttgtgtaa ggcagtgtcg gtgtccatgg gcctgggtgt
1861 ggaggcgttg gagcatctga gtccagggga acagcttcgg ggggattgca tgtacgcccc
1921 acttttggga gttccacccg ctgtgcgtcc cactccttgt gccccattgg ccgaatgcaa
1981 aggttctctg ctagacgaca gcgcaggcaa gagcactgaa gatactgctg agtattcccc
2041 tttcaaggga ggttacacca aagggctaga aggcgagagc ctaggctgct ctggcagcgc

2101 tgcagcaggg agctccggga cacttgaact gccgtctacc ctgtctctct acaagtccgg
2161 agcactggac gaggcagctg cgtaccagag tcgcgactac tacaactttc cactggctct
2221 ggccggaccg ccgccccctc cgccgcctcc ccatccccac gctcgcatca agctggagaa
2281 cccgctggac tacggcagcg cctgggcggc tgcggcggcg cagtgccgct atggggacct
2341 ggcgagcctg catggcgcgg gtgcagcggg acccggttct gggtcaccct cagccgccgc
2401 ttcctcatcc tggcacactc tcttcacagc cgaagaaggc cagttgtatg gaccgtgtgg
2461 tggtggtggg ggtggtggcg gcggcggcgg cggcggcggc ggcggcggcg gcggcggcgg
2521 cggcggcggc gaggcgggag ctgtagcccc ctacggctac actcggcccc ctcaggggct
2581 ggcgggccag gaaagcgact tcaccgcacc tgatgtgtgg taccctggcg gcatggtgag
2641 cagagtgccc tatcccagtc ccacttgtgt caaaagcgaa atgggcccct ggatggatag
2701 ctactccgga ccttacgggg acatgcgttt ggagactgcc agggaccatg ttttgcccat
2761 tgactattac tttccacccc agaagacctg cctgatctgt ggagatgaag cttctgggtg
2821 tcactatgga gctctcacat gtggaagctg caaggtcttc ttcaaaagag ccgctgaagg
2881 gaaacagaag tacctgtgcg ccagcagaaa tgattgcact attgataaat tccgaaggaa
2941 aaattgtcca tcttgtcgtc ttcggaaatg ttatgaagca gggatgactc tgggagcccg
3001 gaagctgaag aaacttggta atctgaaact acaggaggaa ggagaggctt ccagcaccac
3061 cagccccact gaggagacaa cccagaagct gacagtgtca cacattgaag gctatgaatg
3121 tcagcccatc tttctgaatg tcctggaagc cattgagcca ggtgtagtgt gtgctggaca
3181 cgacaacaac cagcccgact cctttgcagc cttgctctct agcctcaatg aactgggaga
3241 gagacagctt gtacacgtgg tcaagtgggc caaggccttg cctggcttcc gcaacttaca
3301 cgtggacgac cagatggctg tcattcagta ctcctggatg gggctcatgg tgtttgccat
3361 gggctggcga tccttcacca atgtcaactc caggatgctc tacttcgccc ctgatctggt
3421 tttcaatgag taccgcatgc acaagtcccg gatgtacagc cagtgtgtcc gaatgaggca
3481 cctctctcaa gagtttggat ggctccaaat caccccccag gaattcctgt gcatgaaagc
3541 actgctactc ttcagcatta ttccagtgga tgggctgaaa aatcaaaaat tctttgatga
3601 acttcgaatg aactacatca aggaactcga tcgtatcatt gcatgcaaaa gaaaaaatcc
3661 cacatcctgc tcaagacgct tctaccagct caccaagctc ctggactccg tgcagcctat
3721 tgcgagagag ctgcatcagt tcacttttga cctgctaatc aagtcacaca tggtgagcgt
3781 ggactttccg gaaatgatgg cagagatcat ctctgtgcaa gtgcccaaga tcctttctgg
3841 gaaagtcaag cccatctatt tccacaccca gtgaagcatt ggaaacccta tttccccacc
3901 ccagctcatg ccccctttca gatgtcttct gcctgttata actctgcact actcctctgc
3961 agtgccttgg ggaatttcct ctattgatgt acagtctgtc atgaacatgt tcctgaattc
4021 tatttgctgg gctttttttt tctctttctc tcctttcttt ttcttcttcc ctccctatct
4081 aaccctccca tggcaccttc agactttgct tcccattgtg gctcctatct gtgttttgaa
4141 tggtgttgta tgcctttaaa tctgtgatga tcctcatatg gcccagtgtc aagttgtgct
4201 tgtttacagc actactctgt gccagccaca caaacgttta cttatcttat gccacgggaa
4261 gtttagagag ctaagattat ctggggaaat caaaacaaaa aacaagcaaa caaaaaaaaa
4321 a

SEQ ID NO:21

1: NM_000044

MEVQLGLGRVYPRPPSKTYRGAFQNLFQSVREVIQNPGPRHPEA
ASAAPPGASLLLLQQQQQQQQQQQQQQQQQQQQQQETSPRQQQQQQGEDGSPQAHRRGP
TGYLVLDEEQQPSQPQSALECHPERGCVPEPGAAVAASKGLPQQLPAPPDEDDSAAPS
TLSLLGPTFPGLSSCSADLKDILSEASTMQLLQQQQQEAVSEGSSSGRAREASGAPTS
SKDNYLGGTSTISDNAKELCKAVSVSMGLGVEALEHLSPGEQLRGDCMYAPLLGVPPA
VRPTPCAPLAECKGSLLDDSAGKSTEDTAEYSPFKGGYTKGLEGESLGCSGSAAAGSS

GTLELPSTLSLYKSGALDEAAAYQSRDYYNFPLALAGPPPPPPPPHPHARIKLENPLD
YGSAWAAAAAQCRYGDLASLHGAGAAGPGSGSPSAAASSSWHTLFTAEEGQLYGPCGG
GGGGGGGGGGGGGGGGGGGGGGEAGAVAPYGYTRPPQGLAGQESDFTAPDVWYPGGMV
SRVPYPSPTCVKSEMGPWMDSYSGPYGDMRLETARDHVLPIDYYFPPQKTCLICGDEA
SGCHYGALTCGSCKVFFKRAAEGKQKYLCASRNDCTIDKFRRKNCPSCRLRKCYEAGM
TLGARKLKKLGNLKLQEEGEASSTTSPTEETTQKLTVSHIEGYECQPIFLNVLEAIEP
GVVCAGHDNNQPDSFAALLSSLNELGERQLVHVVKWAKALPGFRNLHVDDQMAVIQYS
WMGLMVFAMGWRSFTNVNSRMLYFAPDLVFNEYRMHKSRMYSQCVRMRHLSQEFGWLQ
ITPQEFLCMKALLLFSIIPVDGLKNQKFFDELRMNYIKELDRIIACKRKNPTSCSRRF
YQLTKLLDSVQPIARELHQFTFDLLIKSHMVSVDFPEMMAEIISVQVPKILSGKVKPI
YFHTQ

What is claimed is:

1. A method for assessing the risk of prostate cancer in a human subject comprising determining the length of the contiguous CAG or CAA repeats in both AIB1 gene alleles of the subject and assessing whether the length of the CAG or CAA repeats is less than, equal to, or greater than 29 repeats, a length less than or greater than 29 repeats in both alleles indicating an increased risk of prostate cancer in the subject.

2. The method of claim 1, wherein determining the length of the repeats comprises amplifying a region of both AIB1 gene alleles comprising the contiguous CAG or CAA repeat.

3. The method of claim 2, wherein the amplification is by PCR that produces two PCR products.

4. The method of claim 3, further comprising analyzing the PCR products by chromatography.

5. The method of claim 4, wherein the chromatography is gel electrophoresis.

6. The method of claim 2, wherein the sequence of the PCR products is determined.

7. The method of claim 3, wherein the PCR product is produced using a first AIB1 primer that selectively hybridizes with the sequence set forth in SEQ ID NO:3 and a second AIB1 primer that selectively hybridizes with the sequence set forth in SEQ ID NO:4.

8. The method of claim 7, wherein the first AIB1 primer has the sequence set forth in SEQ ID NO:1 and the second AIB1 primer has the sequence set forth in SEQ ID NO:2.

9. The method of claim 1, wherein determining how many CAG or CAA repeats there are comprises sequencing the CAG or CAA repeats.

10. A method for assessing the risk of prostate cancer in a human subject comprising determining the length of the contiguous CAG or CAA repeats in the AIB1 gene alleles of the subject and assessing whether the length of the CAG or CAA repeats in each allele is less than, equal to, or greater than 29 repeats, and determining the length of the contiguous CAG or CAA repeats in the androgen receptor gene of the subject and assessing whether the length of the CAG or CAA repeats is less than, equal to, or greater than 23 repeats, a length in at least one allele less than or greater than 29 repeats in the AIB1 gene and less than or greater than 23 repeats in the androgen receptor gene indicating an increased risk of prostate cancer in the subject.

11. The method of claim 10, wherein determining the length of the contiguous CAG or CAA repeats in the AIB1 gene alleles comprises amplifying a region of the AIB1 gene alleles comprising the CAG or CAA repeats and wherein determining the length of the CAG or CAA repeats in the androgen receptor gene comprises amplifying a region of the androgen receptor gene comprising the CAG or CAA repeats.

12. The method of claim 11, wherein the amplification of the regions of the AIB1 gene and the androgen receptor gene is by PCR that produces a first and a second AIB1 PCR product and an androgen receptor PCR product.

13. The method of claim 12, further comprising analyzing the PCR products by chromatography.

14. The method of claim 13, wherein the chromatography is gel electrophoresis.

15. The method of claim 12, wherein the sequence of the PCR products is determined.

16. The method of claim 12, wherein the AIB1 PCR product is produced using a first AIB1 primer that selectively hybridizes with the sequence set forth in SEQ ID NO:3 and a second AIB1 primer that selectively hybridizes with the sequence set forth in SEQ ID NO:4.

17. The method of claim 16, wherein the first AIB1 primer has the sequence set forth in SEQ ID NO:1 and the second AIB1 primer has the sequence set forth in SEQ ID NO:2.

18. The method of claim 12, wherein the androgen receptor PCR product is produced using a first androgen receptor CAG primer that selectively hybridizes with the sequence set forth in SEQ ID NO:9 and a second androgen receptor CAG primer that selectively hybridizes with the sequence set forth in SEQ ID NO:10.

19. The method of claim 18, wherein the first androgen receptor CAG primer has the sequence set forth in SEQ ID NO:7 and the second androgen receptor CAG primer has the sequence set forth in SEQ ID NO:8.

20. The method of claim 12, wherein the AIB1 PCR product is produced using a first AIB1 primer that selectively hybridizes with the sequence set forth in SEQ ID NO:3 and a second AIB1 primer that selectively hybridizes with the sequence set forth in SEQ ID NO:4. and wherein the androgen receptor PCR product is produced using a first androgen receptor CAG primer that selectively hybridizes with the sequence set forth in SEQ ID NO:9 and a second androgen receptor CAG primer that selectively hybridizes with the sequence set forth in SEQ ID NO:10.

21. The method of claim 20, wherein the first AIB1 primer has the sequence set forth in SEQ ID NO:1 and the second AIB1 primer has the sequence set forth in SEQ ID NO:2 and wherein the first androgen receptor CAG primer has the sequence set forth in SEQ ID NO:7 and the second androgen receptor CAG primer has the sequence set forth in SEQ ID NO:8.

22. The method of claim 10, wherein more than 29 contiguous CAG or CAA repeats in at least one allele of the AIB1 gene of the person and more than 23 contiguous CAG or CAA repeats in the androgen receptor gene of the person indicates an increased risk of prostate cancer in the subject.

23. The method of claim 10, wherein more than 29 contiguous CAG or CAA repeats in at least one allele of the AIB1 gene of the person and less than 23 contiguous CAG or CAA repeats in the androgen receptor gene of the person

indicates an increased risk of prostate cancer in the subject.

24. The method of claim 10, wherein less than 29 contiguous CAG or CAA repeats in at least one allele of the AIB1 gene of the person and more than 23 contiguous CAG or CAA repeats in the androgen receptor gene of the person indicates an increased risk of prostate cancer in the subject.

25. The method of claim 10, wherein less than 29 contiguous CAG or CAA repeats in at least one allele of the AIB1 gene of the person and less than 23 contiguous CAG or CAA repeats in the androgen receptor gene of the person indicates an increased risk of prostate cancer in the subject.

26. The method of claim 10, wherein more or less than 29 contiguous CAG or CAA repeats in both alleles of the AIB1 gene of the person and more than 23 contiguous CAG or CAA repeats in the androgen receptor gene of the person indicates an increased risk of prostate cancer in the subject.

27. The method of claim 10, wherein more or less than 29 contiguous CAG or CAA repeats in both alleles of the AIB1 gene of the person and less than 23 contiguous CAG or CAA repeats in the androgen receptor gene of the person indicates an increased risk of prostate cancer in the subject.

28. The method of claim 10, wherein more than 29 contiguous CAG or CAA repeats in both alleles of the AIB1 gene of the person and more than 23 contiguous CAG or CAA repeats in the androgen receptor gene of the person indicates an increased risk of prostate cancer in the subject.

29. The method of claim 10, wherein less than 29 contiguous CAG or CAA repeats in both alleles of the AIB1 gene of the person and less than 23 contiguous CAG or CAA repeats in the androgen receptor gene of the person indicates an increased risk of prostate cancer in the subject.

30. A method for assessing the risk of prostate cancer in a human subject comprising determining the length of the contiguous CAG or CAA repeats in the AIB1 gene alleles of the subject and assessing whether the length of the CAG or CAA repeats in each allele is less than, equal to, or greater than 29 repeats, and

determining the length of the contiguous GGN repeats in the androgen receptor gene of the subject and assessing whether the length of the GGN repeats is less than, equal to, or greater than 23 repeats, a length in at least one allele less than or greater than 29 repeats in the AIB1 gene and less than or greater than 23 repeats in the androgen receptor gene indicating an increased risk of prostate cancer in the subject, wherein N is either T, G, or C.

31. The method of claim 30, wherein determining the length of the contiguous CAG or CAA repeats in the AIB1 gene alleles comprises amplifying a region of the AIB1 gene alleles comprising the CAG or CAA repeats and wherein determining the length of the contiguous GGN repeats in the androgen receptor gene comprises amplifying a region of the androgen receptor gene comprising the GGN repeats.

32. The method of claim 31, wherein the amplification of the regions of the AIB1 gene and the androgen receptor gene is by PCR that produces a first and a second AIB1 PCR product and an androgen receptor PCR product.

33. The method of claim 32, further comprising analyzing the PCR products by chromatography.

34. The method of claim 33, wherein the chromatography is gel electrophoresis.

35. The method of claim 32, wherein the sequence of the PCR products is determined.

36. The method of claim 32, wherein the AIB1 PCR product is produced using a first AIB1 primer that selectively hybridizes with the sequence set forth in SEQ ID NO:3 and a second AIB1 primer that selectively hybridizes with the sequence set forth in SEQ ID NO:4.

37. The method of claim 36, wherein the first AIB1 primer has the sequence set forth in SEQ ID NO:1 and the second AIB1 primer has the sequence set forth in SEQ ID NO:2.

38. The method of claim 32, wherein the androgen receptor PCR product is produced using a first androgen receptor GGN primer that selectively hybridizes with the sequence set forth in SEQ ID NO:13 and a second androgen receptor GGN primer that selectively hybridizes with the sequence set forth in SEQ ID NO:14.

39. The method of claim 38, wherein the first androgen receptor GGN primer has the sequence set forth in SEQ ID NO:11 and the second androgen receptor GGN primer has the sequence set forth in SEQ ID NO:12.

40. The method of claim 32, wherein the AIB1 PCR product is produced using a first AIB1 primer that selectively hybridizes with the sequence set forth in SEQ ID NO:3 and a second AIB1 primer that selectively hybridizes with the sequence set forth in SEQ ID NO:4. and wherein the androgen receptor PCR product is produced using a first androgen receptor GGN primer that selectively hybridizes with the sequence set forth in SEQ ID NO:13 and a second androgen receptor GGN primer that selectively hybridizes with the sequence set forth in SEQ ID NO:14.

41. The method of claim 40, wherein the first AIB1 primer has the sequence set forth in SEQ ID NO:1 and the second AIB1 primer has the sequence set forth in SEQ ID NO:2 and wherein the first androgen receptor GGN primer has the sequence set forth in SEQ ID NO:11 and the second androgen receptor GGN primer has the sequence set forth in SEQ ID NO:12.

42. The method of claim 30, wherein more than 29 contiguous CAG or CAA repeats in at least one allele of the AIB1 gene of the person and more than 23 contiguous GGN repeats in the androgen receptor gene of the person indicates an increased risk of prostate cancer in the subject.

43. The method of claim 30, wherein more than 29 contiguous CAG or CAA repeats in at least one allele of the AIB1 gene of the person and less than 23 contiguous GGN repeats in the androgen receptor gene of the person indicates an increased risk of prostate cancer in the subject.

44. The method of claim 30, wherein less than 29 contiguous CAG or CAA repeats in at least one allele of the AIB1 gene of the person and more than 23

contiguous GGN repeats in the androgen receptor gene of the person indicates an increased risk of prostate cancer in the subject.

45. The method of claim 30, wherein less than 29 contiguous CAG or CAA repeats in at least one allele of the AIB1 gene of the person and less than 23 contiguous GGN repeats in the androgen receptor gene of the person indicates an increased risk of prostate cancer in the subject.

46. The method of claim 30, wherein more or less than 29 contiguous CAG or CAA repeats in both alleles of the AIB1 gene of the person and more than 23 contiguous GGN repeats in the androgen receptor gene of the person indicates an increased risk of prostate cancer in the subject.

47. The method of claim 30, wherein more or less than 29 contiguous CAG or CAA repeats in both alleles of the AIB1 gene of the person and less than 23 contiguous GGN repeats in the androgen receptor gene of the person indicates an increased risk of prostate cancer in the subject.

48. The method of claim 30, wherein more than 29 contiguous CAG or CAA repeats in both alleles of the AIB1 gene of the person and more than 23 contiguous GGN repeats in the androgen receptor gene of the person indicates an increased risk of prostate cancer in the subject.

49. The method of claim 30, wherein less than 29 contiguous CAG or CAA repeats in both alleles of the AIB1 gene of the person and less than 23 contiguous GGN repeats in the androgen receptor gene of the person indicates an increased risk of prostate cancer in the subject.

50. A method for assessing the risk of prostate cancer in a human subject comprising determining the length of the contiguous CAG or CAA repeats in an AIB1 gene of the subject and assessing whether the length of the CAG repeats is less than, equal to, or greater than 29 repeats, a length less than or greater than 29 repeats indicating an increased risk of prostate cancer in the subject.

51. A method for assessing the risk of prostate cancer in a human subject comprising determining the length of the contiguous CAG or CAA repeats in an

AIB1 gene of the subject and assessing whether the length of the CAG or CAA repeats is less than, equal to, or greater than 29 repeats, and determining the length of the contiguous CAG or CAA repeats in the androgen receptor gene of the subject and assessing whether the length of the CAG or CAA repeats is less than, equal to, or greater than 23 repeats, a length less than or greater than 29 repeats in the AIB1 allele and less than or greater than 23 repeats in the androgen receptor gene indicating an increased risk of prostate cancer in the subject.

52. A method for assessing the risk of prostate cancer in a human subject comprising determining the length of the contiguous CAG or CAA repeats in an AIB1 gene of the subject and assessing whether the length of the CAG or CAA repeats is less than, equal to, or greater than 29 repeats, and determining the length of the contiguous GGN repeats in the androgen receptor gene of the subject and assessing whether the length of the GGN repeats is less than, equal to, or greater than 23 repeats, a length less than or greater than 29 repeats in the AIB1 allele and less than or greater than 23 repeats in the androgen receptor gene indicating an increased risk of prostate cancer in the subject, wherein N is either T, G, or C.

53. A kit for assessing a subject's risk for acquiring prostate cancer, comprising the oligonucleotides set forth in SEQ ID Nos: 1 and 2.

54. A composition comprising the primers having the sequence set forth in SEQ ID Nos: 1 and 2.

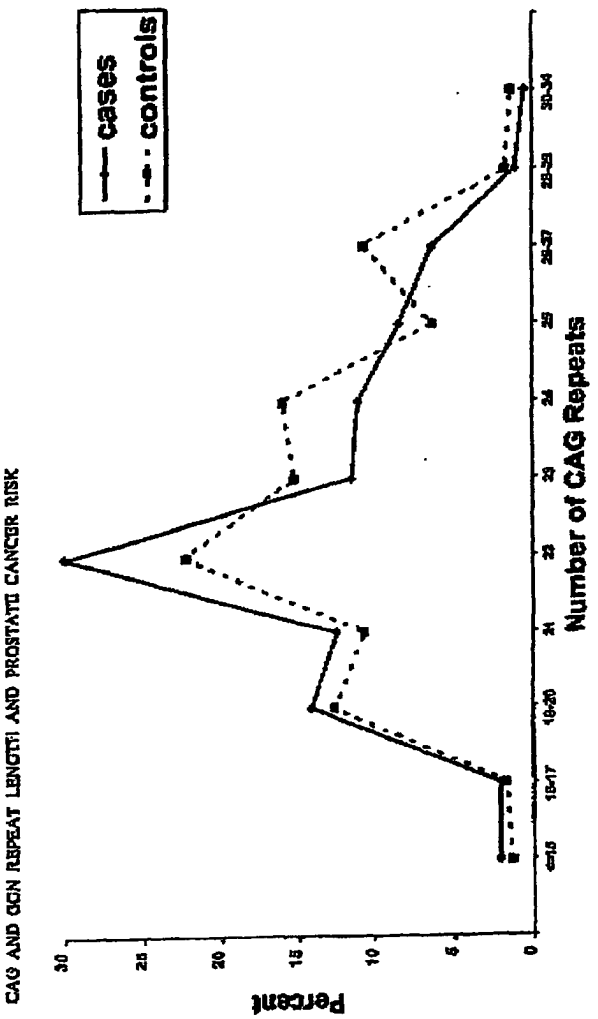

Fig. 1. Percent distribution of number of CAG repeats.

FIG. 1

SEQUENCE LISTING

<110> Chang, Chawnshang
Hsing, Ann

<120> METHODS AND COMPOSITIONS FOR PREDICTING PROSTATE CANCER

<130> 21108.0001U1

<160> 21

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note = synthetic construct

<221> misc_feature
<222> (0)...(0)

<400> 1
tcatcacctc cgacaacaga gg 22

<210> 2
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note = synthetic construct

<400> 2
tatggaaact gttgcggagg ag 22

<210> 3
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note = synthetic construct

<400> 3
cctctgttgt cggaggtgat ga 22

<210> 4
<211> 22
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 4
ctcctccgca acagtttcca ta 22

<210> 5
<211> 48
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> 1, 2, 3, 7, 8, 9
<223> Sequence can be repeated one or more times

<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 5
cagcaacagc aacagcaaca gcaacagcaa cagcagcaac agcagcaa 48

<210> 6
<211> 87
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 6
cagcagcagc agcagcagca acagcagcag cagcagcagc agcagcagca acagcaacag 60
caacagcaac agcagcaaca gcagcaa 87

<210> 7
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 7
gctctgggac gcaacctctc t 21

<210> 8
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 8
gcagcgacta ccgcatcatc a 21

<210> 9

<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 9
agagaggttg cgtcccagag c 21

<210> 10
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 10
tgatgatgcg gtagtcgctg c 21

<210> 11
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 11
accctcagcc gccgcttcct catc 24

<210> 12
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 12
ctgggatagg gcactctgct caac 24

<210> 13
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 13
gatgaggaag cggcggctga gggt 24

<210> 14

<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 14
gttgagcaga gtgccctatc ccag 24

<210> 15
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 15
cggggtaagg gaagtaggtg gaag 24

<210> 16
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 16
ctctacgatg ggcttgggga gaac 24

<210> 17
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<221> misc_feature
<222> 21
<223> Sequence can be repeated one or more times

<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 17
ggtggtggtg ggggtggtgg c 21

<210> 18
<211> 6832
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 18
cggcagcggc tgcggcttag tcggtggcgg ccggcggcgg ctgcgggctg agcggcgagt 60
ttccgattta aagctgagct gcgaggaaaa tggcggcggg aggatcaaaa tacttgctgg 120
atggtggact cagagaccaa taaaaataaa ctgcttgaac atcctttgac tggttagcca 180
gttgctgatg tatattcaag atgagtggat taggagaaaa cttggatcca ctggccagtg 240
attcacgaaa acgcaaattg ccatgtgata ctccaggaca aggtcttacc tgcagtggtg 300
aaaaacggag acgggagcag gaaagtaaat atattgaaga attggctgag ctgatatctg 360
ccaatcttag tgatattgac aatttcaatg tcaaaccaga taaatgtgcg attttaaagg 420
aaacagtaag acagatacgt caaataaaag agcaaggaaa aactatttcc aatgatgatg 480
atgttcaaaa agccgatgta tcttctacag ggcagggagt tattgataaa gactccttag 540
gaccgctttt acttcaggca ttggatggtt tcctatttgt ggtgaatcga gacggaaaca 600
ttgtatttgt atcagaaaat gtcacacaat acctgcaata taagcaagag gacctggtta 660
acacaagtgt ttacaatatc ttacatgaag aagacagaaa ggattttctt aagaatttac 720
caaaatctac agttaatgga gtttcctgga caaatgagac ccaaagacaa aaaagccata 780
catttaattg ccgtatgttg atgaaaacac cacatgatat tctggaagac ataaacgcca 840
gtcctgaaat gcgccagaga tatgaaacaa tgcagtgctt tgccctgtct cagccacgag 900
ctatgatgga ggaaggggaa gatttgcaat cttgtatgat ctgtgtggca cgccgcatta 960
ctacaggaga aagaacattt ccatcaaacc ctgagagctt tattaccaga catgatcttt 1020
caggaaaggt tgtcaatata gatacaaatt cactgagatc ctccatgagg cctggctttg 1080
aagatataat ccgaaggtgt attcagagat tttttagtct aaatgatggg cagtcatggt 1140
cccagaaacg tcactatcaa gaagcttatc ttaatggcca tgcagaaacc ccagtatatc 1200
gattctcgtt ggctgatgga actatagtga ctgcacagac aaaaagcaaa ctcttccgaa 1260
atcctgtaac aaatgatcga catggctttg tctcaaccca cttccttcag agagaacaga 1320
atggatatag accaaaccca aatcctgttg gacaagggat tagaccacct atggctggat 1380
gcaacagttc ggtaggcggc atgagtatgt cgccaaacca aggcttacag atgccgagca 1440
gcagggccta tggcttggca gaccctagca ccacagggca gatgagtgga gctaggtatg 1500
ggggttccag taacatagct tcattgaccc ctgggccagg catgcaatca ccatcttcct 1560
accagaacaa caactatggg ctcaacatga gtagcccccc acatgggagt cctggtcttg 1620
ccccaaacca gcagaatatc atgatttctc ctcgtaatcg tgggagtcca aagatagcct 1680
cacatcagtt ttctcctgtt gcaggtgtgc actctcccat ggcatcttct ggcaatactg 1740
ggaaccacag cttttccagc agctctctca gtgccctgca agccatcagt gaaggtgtgg 1800
ggacttccct tttatctact ctgtcatcac caggccccaa attggataac tctcccaata 1860
tgaatattac ccaaccaagt aaagtaagca atcaggattc caagagtcct ctgggctttt 1920
attgcgacca aaatccagtg gagagttcaa tgtgtcagtc aaatagcaga gatcacctca 1980
gtgacaaaga aagtaaggag agcagtgttg aggggcaga gaatcaaagg ggtcctttgg 2040
aaagcaaagg tcataaaaaa ttactgcagt tacttacctg ttcttctgat gaccggggtc 2100
attcctcctt gaccaactcc cccctagatt caagttgtaa agaatcttct gttagtgtca 2160
ccagcccctc tggagtctcc tcctctacat ctggaggagt atcctctaca tccaatatgc 2220
atgggtcact gttacaagag aagcaccgga ttttgcacaa gttgctgcag aatgggaatt 2280
caccagctga ggtagccaag attactgcag aagccactgg gaaagacacc agcagtataa 2340
cttcttgtgg ggacggaaat gttgtcaagc aggagcagct aagtcctaag aagaaggaga 2400
ataatgcact tcttagatac ctgctggaca gggatgatcc tagtgatgca ctctctaaag 2460
aactacagcc ccaagtggaa ggagtggata ataaaatgag tcagtgcacc agctccacca 2520
ttcctagctc aagtcaagag aaagacccta aaattaagac agagacaagt gaagagggat 2580
ctggagactt ggataatcta gatgctattc ttggtgatct gactagttct gacttttaca 2640
ataattccat atcctcaaat ggtagtcatc tggggactaa gcaacaggtg tttcaaggaa 2700
ctaattctct gggtttgaaa agttcacagt ctgtgcagtc tattcgtcct ccatataacc 2760
gagcagtgtc tctggatagc cctgtttctg ttggctcaag tcctccagta aaaaatatca 2820
gtgctttccc catgttacca aagcaaccca tgttgggtgg gaatccaaga atgatggata 2880
gtcaggaaaa ttatggctca agtatgggtg ggccaaaccg aaatgtgact gtgactcaga 2940
ctccttcctc aggagactgg ggcttaccaa actcaaaggc cggcagaatg gaacctatga 3000
attcaaactc catgggaaga ccaggaggag attataatac ttctttaccc agacctgcac 3060
tgggtggctc tattcccaca ttgcctcttc ggtctaatag cataccaggt gcgagaccag 3120
tattgcaaca gcagcagcag atgcttcaaa tgaggcctgg tgaaatcccc atgggaatgg 3180
gggctaatcc ctatggccaa gcagcagcat ctaaccaact gggttcctgg cccgatggca 3240
tgttgtccat ggaacaagtt tctcatggca ctcaaaatag gcctcttctt aggaattccc 3300
tggatgatct tgttgggcca ccttccaacc tggaaggcca gagtgacgaa agagcattat 3360
tggaccagct gcacactctt ctcagcaaca cagatgccac aggcctggaa gaaattgaca 3420
gagctttggg cattcctgaa cttgtcaatc agggacaggc attagagccc aaacaggatg 3480
ctttccaagg ccaagaagca gcagtaatga tggatcagaa ggcaggatta tatggacaga 3540

catacccagc acaggggcct ccaatgcaag gaggctttca tcttcaggga caatcaccat 3600
cttttaactc tatgatgaat cagatgaacc agcaaggcaa ttttcctctc caaggaatgc 3660
acccacgagc caacatcatg agaccccgga caaacacccc caagcaactt agaatgcagc 3720
ttcagcagag gctgcagggc cagcagtttt tgaatcagag ccgacaggca cttgaattga 3780
aaatggaaaa ccctactgct ggtggtgctg cggtgatgag gcctatgatg cagccccagc 3840
agggttttct taatgctcaa atggtcgccc aacgcagcag agagctgcta agtcatcact 3900
tccgacaaca gagggtggct atgatgatgc agcagcagca gcagcagcaa cagcagcagc 3960
agcagcagca gcagcagcaa cagcaacagc aacagcaaca gcagcaacag cagcaaaccc 4020
aggccttcag cccacctcct aatgtgactg cttcccccag catggatggg cttttggcag 4080
gacccacaat gccacaagct cctccgcaac agtttccata tcaaccaaat tatggaatgg 4140
gacaacaacc agatccagcc tttggtcgag tgtctagtcc tcccaatgca atgatgtcgt 4200
caagaatggg tccctcccag aatcccatga tgcaacaccc gcaggctgca tccatctatc 4260
agtcctcaga aatgaagggc tggccatcag gaaatttggc caggaacagc tccttttccc 4320
agcagcagtt tgcccaccag gggaatcctg cagtgtatag tatggtgcac atgaatggca 4380
gcagtggtca catgggacag atgaacatga accccatgcc catgtctggc atgcctatgg 4440
gtcctgatca gaatactgct gacatctctg caccaggacc tcttaaggaa accactgtac 4500
aaatgacact gcactaggat tattgggaag gaatcattgt tccaggcatc catcttggaa 4560
gaaaggacca gctttgagct ccatcaaggg tattttaagt gatgtcattt gagcaggact 4620
ggattttaag ccgaagggca atatctacgt gtttttcccc cctccttctg ctgtgtatca 4680
tggtgttcaa aacagaaatg ttttttggca ttccacctcc tagggatata attctggaga 4740
catggagtgt tactgatcat aaaacttttg tgtcactttt ttctgccttg ctagccaaaa 4800
tctcttaaat acacgtaggt gggccagaga acattggaag aatcaagaga gattagaata 4860
tctggtttct ctagttgcag tattggacaa agagcatagt cccagccttc aggtgtagta 4920
gttctgtgtt gacccttttgt ccagtggaat tggtgattct gaattgtcct ttactaatgg 4980
tgttgagttg ctctgtccct attatttgcc ctaggctttc tcctaatgaa ggttttcatt 5040
tgccattcat gtcctgtaat acttcacctc caggaactgt catggatgtc caaatggctt 5100
tgcagaaagg aaatgagatg acagtattta atcgcagcag tagcaaactt ttcacatgct 5160
aatgtgcagc tgagtgcact ttatttaaaa agaatggata aatgcaatat tcttgaggtc 5220
ttgagggaat agtgaaacac attcctggtt tttgcctaca cttacgtgtt agacaagaac 5280
tatgattttt ttttttaaag tactggtgtc acccttttgcc tatatggtag agcaataatg 5340
ctttttaaaa ataaacttct gaaaacccaa ggccaggtac tgcattctga atcagaatct 5400
cgcagtgttt ctgtgaatag attttttttgt aaatatgacc tttaagatat tgtattatgt 5460
aaaatatgta tatacctttt tttgtaggtc acaacaactc atttttacag agtttgtgaa 5520
gctaaatatt taacattgtt gatttcagta agctgtgtgg tgaggctacc agtggaagag 5580
acatcccttg acttttgtgg cctgggggag gggtagtgct ccacagcttt tccttcccca 5640
ccccccagcc ttagatgcct cgctcttttc aatctcttaa tctaaatgct ttttaaagag 5700
attatttgtt tagatgtagg cattttaatt ttttaaaaat tcctctacca gaactaagca 5760
ctttgttaat ttgggggggaa agaatagata tggggaaata aacttaaaaa aaaatcagga 5820
atttaaaaaa acgagcaatt tgaagagaat cttttggatt ttaagcagtc cgaaataata 5880
gcaattcatg ggctgtgtgt gtgtgtgtat gtgtgtgtgt gtgtgtgtat gtttaattat 5940
gttacctttt catccccttt aggagcgttt tcagattttg gttgctaaga cctgaatccc 6000
atattgagat ctcgagtaga atccttggtg tggtttctgg tgtctgctca gctgtcccct 6060
cattctacta atgtgatgct ttcattatgt ccctgtggat tagaatagtg tcagttattt 6120
cttaagtaac tcagtaccca gaacagccag ttttactgtg attcagagcc acagtctaac 6180
tgagcacctt ttaaacccct ccctcttctg ccccctacca cttttctgct gttgcctctc 6240
tttgacacct gttttagtca gttgggagga agggaaaaat caagtttaat tccctttatc 6300
tgggttaatt catttggttc aaatagttga cggaattggg tttctgaatg tctgtgaatt 6360
tcagaggtct ctgctagcct tggtatcatt ttctagcaat aactgagagc cagttaattt 6420
taagaatttc acacatttag ccaatctttc tagatgtctc tgaaggtaag atcatttaat 6480
atctttgata tgcttacgag taagtgaatc ctgattattt ccagacccac caccagagtg 6540
gatcttattt tcaaagcagt atagacaatt atgagtttgc cctctttccc ctaccaagtt 6600
caaaatatat ctaagaaaga ttgtaaatcc gaaaacttcc attgtagtgg cctgtgcttt 6660
tcagatagta tactctcctg tttggagaca gaggaagaac caggtcagtc tgtctctttt 6720
tcagctcaat tgtatctgac ccttctttaa gttatgtgtg tggggagaaa tagaatggtg 6780
ctcttatctt tcttgacttt aaaaaaatta ttaaaaacaa aaaaaaaata aa 6832

<210> 19
<211> 1438

<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 19
Met Ser Gly Leu Gly Glu Asn Leu Asp Pro Leu Ala Ser Asp Ser Arg
1 5 10 15
Lys Arg Lys Leu Pro Cys Asp Thr Pro Gly Gln Gly Leu Thr Cys Ser
20 25 30
Gly Glu Lys Arg Arg Arg Glu Gln Glu Ser Lys Tyr Ile Glu Glu Leu
35 40 45
Ala Glu Leu Ile Ser Ala Asn Leu Ser Asp Ile Asp Asn Phe Asn Val
50 55 60
Lys Pro Asp Lys Cys Ala Ile Leu Lys Glu Thr Val Arg Gln Ile Arg
65 70 75 80
Gln Ile Lys Glu Gln Gly Lys Thr Ile Ser Asn Asp Asp Asp Val Gln
85 90 95
Lys Ala Asp Val Ser Ser Thr Gly Gln Gly Val Ile Asp Lys Asp Ser
100 105 110
Leu Gly Pro Leu Leu Leu Gln Ala Leu Asp Gly Phe Leu Phe Val Val
115 120 125
Asn Arg Asp Gly Asn Ile Val Phe Val Ser Glu Asn Val Thr Gln Tyr
130 135 140
Leu Gln Tyr Lys Gln Glu Asp Leu Val Asn Thr Ser Val Tyr Asn Ile
145 150 155 160
Leu His Glu Glu Asp Arg Lys Asp Phe Leu Lys Asn Leu Pro Lys Ser
165 170 175
Thr Val Asn Gly Val Ser Trp Thr Asn Glu Thr Gln Arg Gln Lys Ser
180 185 190
His Thr Phe Asn Cys Arg Met Leu Met Lys Thr Pro His Asp Ile Leu
195 200 205
Glu Asp Ile Asn Ala Ser Pro Glu Met Arg Gln Arg Tyr Glu Thr Met
210 215 220
Gln Cys Phe Ala Leu Ser Gln Pro Arg Ala Met Met Glu Glu Gly Glu
225 230 235 240
Asp Leu Gln Ser Cys Met Ile Cys Val Ala Arg Arg Ile Thr Thr Gly
245 250 255
Glu Arg Thr Phe Pro Ser Asn Pro Glu Ser Phe Ile Thr Arg His Asp
260 265 270
Leu Ser Gly Lys Val Val Asn Ile Asp Thr Asn Ser Leu Arg Ser Ser
275 280 285
Met Arg Pro Gly Phe Glu Asp Ile Ile Arg Arg Cys Ile Gln Arg Phe
290 295 300
Phe Ser Leu Asn Asp Gly Gln Ser Trp Ser Gln Lys Arg His Tyr Gln
305 310 315 320
Glu Ala Tyr Leu Asn Gly His Ala Glu Thr Pro Val Tyr Arg Phe Ser
325 330 335
Leu Ala Asp Gly Thr Ile Val Thr Ala Gln Thr Lys Ser Lys Leu Phe
340 345 350
Arg Asn Pro Val Thr Asn Asp Arg His Gly Phe Val Ser Thr His Phe
355 360 365
Leu Gln Arg Glu Gln Asn Gly Tyr Arg Pro Asn Pro Asn Pro Val Gly
370 375 380
Gln Gly Ile Arg Pro Pro Met Ala Gly Cys Asn Ser Ser Val Gly Gly
385 390 395 400
Met Ser Met Ser Pro Asn Gln Gly Leu Gln Met Pro Ser Ser Arg Ala
405 410 415

```
Tyr Gly Leu Ala Asp Pro Ser Thr Thr Gly Gln Met Ser Gly Ala Arg
            420                 425                 430
Tyr Gly Gly Ser Ser Asn Ile Ala Ser Leu Thr Pro Gly Pro Gly Met
        435                 440                 445
Gln Ser Pro Ser Ser Tyr Gln Asn Asn Asn Tyr Gly Leu Asn Met Ser
    450                 455                 460
Ser Pro Pro His Gly Ser Pro Gly Leu Ala Pro Asn Gln Gln Asn Ile
465                 470                 475                 480
Met Ile Ser Pro Arg Asn Arg Gly Ser Pro Lys Ile Ala Ser His Gln
                485                 490                 495
Phe Ser Pro Val Ala Gly Val His Ser Pro Met Ala Ser Ser Gly Asn
            500                 505                 510
Thr Gly Asn His Ser Phe Ser Ser Ser Ser Leu Ser Ala Leu Gln Ala
        515                 520                 525
Ile Ser Glu Gly Val Gly Thr Ser Leu Leu Ser Thr Leu Ser Ser Pro
    530                 535                 540
Gly Pro Lys Leu Asp Asn Ser Pro Asn Met Asn Ile Thr Gln Pro Ser
545                 550                 555                 560
Lys Val Ser Asn Gln Asp Ser Lys Ser Pro Leu Gly Phe Tyr Cys Asp
                565                 570                 575
Gln Asn Pro Val Glu Ser Ser Met Cys Gln Ser Asn Ser Arg Asp His
            580                 585                 590
Leu Ser Asp Lys Glu Ser Lys Glu Ser Ser Val Glu Gly Ala Glu Asn
        595                 600                 605
Gln Arg Gly Pro Leu Glu Ser Lys Gly His Lys Lys Leu Leu Gln Leu
    610                 615                 620
Leu Thr Cys Ser Ser Asp Asp Arg Gly His Ser Ser Leu Thr Asn Ser
625                 630                 635                 640
Pro Leu Asp Ser Ser Cys Lys Glu Ser Ser Val Ser Val Thr Ser Pro
                645                 650                 655
Ser Gly Val Ser Ser Ser Thr Ser Gly Gly Val Ser Ser Thr Ser Asn
            660                 665                 670
Met His Gly Ser Leu Leu Gln Glu Lys His Arg Ile Leu His Lys Leu
        675                 680                 685
Leu Gln Asn Gly Asn Ser Pro Ala Glu Val Ala Lys Ile Thr Ala Glu
    690                 695                 700
Ala Thr Gly Lys Asp Thr Ser Ser Ile Thr Ser Cys Gly Asp Gly Asn
705                 710                 715                 720
Val Val Lys Gln Glu Gln Leu Ser Pro Lys Lys Lys Glu Asn Asn Ala
                725                 730                 735
Leu Leu Arg Tyr Leu Leu Asp Arg Asp Asp Pro Ser Asp Ala Leu Ser
            740                 745                 750
Lys Glu Leu Gln Pro Gln Val Glu Gly Val Asp Asn Lys Met Ser Gln
        755                 760                 765
Cys Thr Ser Ser Thr Ile Pro Ser Ser Ser Gln Glu Lys Asp Pro Lys
    770                 775                 780
Ile Lys Thr Glu Thr Ser Glu Glu Gly Ser Gly Asp Leu Asp Asn Leu
785                 790                 795                 800
Asp Ala Ile Leu Gly Asp Leu Thr Ser Ser Asp Phe Tyr Asn Asn Ser
                805                 810                 815
Ile Ser Ser Asn Gly Ser His Leu Gly Thr Lys Gln Gln Val Phe Gln
            820                 825                 830
Gly Thr Asn Ser Leu Gly Leu Lys Ser Ser Gln Ser Val Gln Ser Ile
        835                 840                 845
Arg Pro Pro Tyr Asn Arg Ala Val Ser Leu Asp Ser Pro Val Ser Val
    850                 855                 860
Gly Ser Ser Pro Pro Val Lys Asn Ile Ser Ala Phe Pro Met Leu Pro
865                 870                 875                 880
Lys Gln Pro Met Leu Gly Gly Asn Pro Arg Met Met Asp Ser Gln Glu
                885                 890                 895
```

```
Asn Tyr Gly Ser Ser Met Gly Gly Pro Asn Arg Asn Val Thr Val Thr
            900                 905                 910
Gln Thr Pro Ser Ser Gly Asp Trp Gly Leu Pro Asn Ser Lys Ala Gly
        915                 920                 925
Arg Met Glu Pro Met Asn Ser Asn Ser Met Gly Arg Pro Gly Gly Asp
    930                 935                 940
Tyr Asn Thr Ser Leu Pro Arg Pro Ala Leu Gly Gly Ser Ile Pro Thr
945                 950                 955                 960
Leu Pro Leu Arg Ser Asn Ser Ile Pro Gly Ala Arg Pro Val Leu Gln
                965                 970                 975
Gln Gln Gln Gln Met Leu Gln Met Arg Pro Gly Glu Ile Pro Met Gly
            980                 985                 990
Met Gly Ala Asn Pro Tyr Gly Gln Ala Ala Ala Ser Asn Gln Leu Gly
        995                 1000                1005
Ser Trp Pro Asp Gly Met Leu Ser Met Glu Gln Val Ser His Gly Thr
    1010                1015                1020
Gln Asn Arg Pro Leu Leu Arg Asn Ser Leu Asp Asp Leu Val Gly Pro
1025                1030                1035                1040
Pro Ser Asn Leu Glu Gly Gln Ser Asp Glu Arg Ala Leu Leu Asp Gln
                1045                1050                1055
Leu His Thr Leu Leu Ser Asn Thr Asp Ala Thr Gly Leu Glu Glu Ile
            1060                1065                1070
Asp Arg Ala Leu Gly Ile Pro Glu Leu Val Asn Gln Gly Gln Ala Leu
        1075                1080                1085
Glu Pro Lys Gln Asp Ala Phe Gln Gly Gln Glu Ala Ala Val Met Met
    1090                1095                1100
Asp Gln Lys Ala Gly Leu Tyr Gly Gln Thr Tyr Pro Ala Gln Gly Pro
1105                1110                1115                1120
Pro Met Gln Gly Gly Phe His Leu Gln Gly Gln Ser Pro Ser Phe Asn
                1125                1130                1135
Ser Met Met Asn Gln Met Asn Gln Gln Gly Asn Phe Pro Leu Gln Gly
            1140                1145                1150
Met His Pro Arg Ala Asn Ile Met Arg Pro Arg Thr Asn Thr Pro Lys
        1155                1160                1165
Gln Leu Arg Met Gln Leu Gln Gln Arg Leu Gln Gly Gln Gln Phe Leu
    1170                1175                1180
Asn Gln Ser Arg Gln Ala Leu Glu Leu Lys Met Glu Asn Pro Thr Ala
1185                1190                1195                1200
Gly Gly Ala Ala Val Met Arg Pro Met Met Gln Pro Gln Gln Gly Phe
                1205                1210                1215
Leu Asn Ala Gln Met Val Ala Gln Arg Ser Arg Glu Leu Leu Ser His
            1220                1225                1230
His Phe Arg Gln Gln Arg Val Ala Met Met Met Gln Gln Gln Gln Gln
        1235                1240                1245
Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln
    1250                1255                1260
Gln Gln Gln Gln Gln Gln Gln Gln Thr Gln Ala Phe Ser Pro Pro Pro
1265                1270                1275                1280
Asn Val Thr Ala Ser Pro Ser Met Asp Gly Leu Leu Ala Gly Pro Thr
                1285                1290                1295
Met Pro Gln Ala Pro Pro Gln Gln Phe Pro Tyr Gln Pro Asn Tyr Gly
            1300                1305                1310
Met Gly Gln Gln Pro Asp Pro Ala Phe Gly Arg Val Ser Ser Pro Pro
        1315                1320                1325
Asn Ala Met Met Ser Ser Arg Met Gly Pro Ser Gln Asn Pro Met Met
    1330                1335                1340
Gln His Pro Gln Ala Ala Ser Ile Tyr Gln Ser Ser Glu Met Lys Gly
1345                1350                1355                1360
Trp Pro Ser Gly Asn Leu Ala Arg Asn Ser Ser Phe Ser Gln Gln Gln
                1365                1370                1375
```

Phe Ala His Gln Gly Asn Pro Ala Val Tyr Ser Met Val His Met Asn
1380 1385 1390
Gly Ser Ser Gly His Met Gly Gln Met Asn Met Asn Pro Met Pro Met
1395 1400 1405
Ser Gly Met Pro Met Gly Pro Asp Gln Asn Thr Ala Asp Ile Ser Ala
1410 1415 1420
Pro Gly Pro Leu Lys Glu Thr Thr Val Gln Met Thr Leu His
1425 1430 1435

<210> 20
<211> 4321
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 20
cgagatcccg gggagccagc ttgctgggag agcgggacgg tccggagcaa gcccacaggc 60
agaggaggcg acagagggaa aaagggccga gctagccgct ccagtgctgt acaggagccg 120
aagggacgca ccacgccagc cccagcccgg ctccagcgac agccaacgcc tcttgcagcg 180
cggcggcttc gaagccgccg cccggagctg ccctttcctc ttcggtgaag tttttaaaag 240
ctgctaaaga ctcggaggaa gcaaggaaag tgcctggtag gactgacggc tgcctttgtc 300
ctcctcctct ccacccccgcc tccccccacc ctgccttccc cccctccccc gtcttctctc 360
ccgcagctgc ctcagtcggc tactctcagc caacccccct caccaccctt ctccccaccc 420
gccccccccgc ccccgtcggc ccagcgctgc cagcccgagt ttgcagagag gtaactccct 480
ttggctgcga gcgggcgagc tagctgcaca ttgcaaagaa ggctcttagg agccaggcga 540
ctggggagcg gcttcagcac tgcagccacg acccgcctgg ttagaattcc ggcggagaga 600
accctctgtt ttcccccact ctctctccac ctcctcctgc cttccccacc ccgagtgcgg 660
agcagagatc aaaagatgaa aaggcagtca ggtcttcagt agccaaaaaa caaaacaaac 720
aaaaacaaaa aagccgaaat aaaagaaaaa gataataact cagttcttat ttgcacctac 780
ttcagtggac actgaatttg gaaggtggag gattttgttt ttttctttta agatctgggc 840
atcttttgaa tctacccttc aagtattaag agacagactg tgagcctagc agggcagatc 900
ttgtccaccg tgtgtcttct tctgcacgag actttgaggc tgtcagagcg ctttttgcgt 960
ggttgctccc gcaagtttcc ttctctggag cttcccgcag gtgggcagct agctgcagcg 1020
actaccgcat catcacagcc tgttgaactc ttctgagcaa gagaagggga ggcggggtaa 1080
gggaagtagg tggaagattc agccaagctc aaggatggaa gtgcagttag ggctgggaag 1140
ggtctaccct cggccgccgt ccaagaccta ccgaggagct ttccagaatc tgttccagag 1200
cgtgcgcgaa gtgatccaga acccgggccc caggcaccca gaggccgcga gcgcagcacc 1260
tcccggcgcc agtttgctgc tgctgcagca gcagcagcag cagcagcagc agcagcagca 1320
gcagcagcag cagcagcagc agcagcaaga gactagcccc aggcagcagc agcagcagca 1380
gggtgaggat ggttctcccc aagcccatcg tagaggcccc acaggctacc tggtcctgga 1440
tgaggaacag caaccttcac agccgcagtc ggccctggag tgccaccccg agagaggttg 1500
cgtcccagag cctggagccg ccgtggccgc cagcaagggg ctgccgcagc agctgccagc 1560
acctccggac gaggatgact cagctgcccc atccacgttg tccctgctgg gccccacttt 1620
ccccggctta agcagctgct ccgctgacct taaagacatc ctgagcgagg ccagcaccat 1680
gcaactcctt cagcaacagc agcaggaagc agtatccgaa ggcagcagca gcgggagagc 1740
gagggaggcc tcgggggctc ccacttcctc caaggacaat tacttagggg gcacttcgac 1800
catttctgac aacgccaagg agttgtgtaa ggcagtgtcg gtgtccatgg gcctgggtgt 1860
ggaggcgttg gagcatctga gtccagggga acagcttcgg ggggattgca tgtacgcccc 1920
acttttggga gttccacccg ctgtgcgtcc cactccttgt gccccattgg ccgaatgcaa 1980
aggttctctg ctagacgaca gcgcaggcaa gagcactgaa gatactgctg agtattcccc 2040
tttcaaggga ggttacacca aagggctaga aggcgagagc ctaggctgct ctggcagcgc 2100
tgcagcaggg agctccggga cacttgaact gccgtctacc ctgtctctct acaagtccgg 2160
agcactggac gaggcagctg cgtaccagag tcgcgactac tacaactttc cactggctct 2220
ggccggaccg ccgcccccctc cgccgcctcc ccatccccac gctcgcatca agctggagaa 2280
cccgctggac tacggcagcg cctgggcggc tgcggcggcg cagtgccgct atggggacct 2340
ggcgagcctg catggcgcgg gtgcagcggg acccggttct gggtcaccct cagccgccgc 2400
ttcctcatcc tggcacactc tcttcacagc cgaagaaggc cagttgtatg gaccgtgtgg 2460

tggtggtggg ggtggtggcg gcggcggcgg cggcggcggc ggcggcggcg gcggcggcgg 2520
cggcggcggc gaggcgggag ctgtagcccc ctacggctac actcggcccc ctcaggggct 2580
ggcgggccag gaaagcgact tcaccgcacc tgatgtgtgg taccctggcg gcatggtgag 2640
cagagtgccc tatcccagtc ccacttgtgt caaaagcgaa atgggccccc ggatggatag 2700
ctactccgga ccttacgggg acatgcgttt ggagactgcc agggaccatg ttttgcccat 2760
tgactattac tttccacccc agaagacctg cctgatctgt ggagatgaag cttctgggtg 2820
tcactatgga gctctcacat gtggaagctg caaggtcttc ttcaaaagag ccgctgaagg 2880
gaaacagaag tacctgtgcg ccagcagaaa tgattgcact attgataaat tccgaaggaa 2940
aaattgtcca tcttgtcgtc ttcggaaatg ttatgaagca gggatgactc tgggagcccg 3000
gaagctgaag aaacttggta atctgaaact acaggaggaa ggagaggctt ccagcaccac 3060
cagccccact gaggagacaa cccagaagct gacagtgtca cacattgaag gctatgaatg 3120
tcagcccatc tttctgaatg tcctggaagc cattgagcca ggtgtagtgt gtgctggaca 3180
cgacaacaac cagcccgact cctttgcagc cttgctctct agcctcaatg aactgggaga 3240
gagacagctt gtacacgtgg tcaagtgggc caaggccttg cctggcttcc gcaacttaca 3300
cgtggacgac cagatggctg tcattcagta ctcctggatg gggctcatgg tgtttgccat 3360
gggctggcga tccttcacca atgtcaactc caggatgctc tacttcgccc ctgatctggt 3420
tttcaatgag taccgcatgc acaagtcccg gatgtacagc cagtgtgtcc gaatgaggca 3480
cctctctcaa gagtttggat ggctccaaat caccccccag gaattcctgt gcatgaaagc 3540
actgctactc ttcagcatta ttccagtgga tgggctgaaa aatcaaaaat tctttgatga 3600
acttcgaatg aactacatca aggaactcga tcgtatcatt gcatgcaaaa gaaaaaatcc 3660
cacatcctgc tcaagacgct tctaccagct caccaagctc ctggactccg tgcagcctat 3720
tgcgagagag ctgcatcagt tcacttttga cctgctaatc aagtcacaca tggtgagcgt 3780
ggactttccg gaaatgatgg cagagatcat ctctgtgcaa gtgcccaaga tcctttctgg 3840
gaaagtcaag cccatctatt tccacaccca gtgaagcatt ggaaacccta tttccccacc 3900
ccagctcatg ccccctttca gatgtcttct gcctgttata actctgcact actcctctgc 3960
agtgccttgg ggaatttcct ctattgatgt acagtctgtc atgaacatgt tcctgaattc 4020
tatttgctgg gctttttttt tctctttctc tcctttcttt ttcttcttcc ctccctatct 4080
aaccctccca tggcaccttc agactttgct tcccattgtg gctcctatct gtgttttgaa 4140
tggtgttgta tgcctttaaa tctgtgatga tcctcatatg gcccagtgtc aagttgtgct 4200
tgtttacagc actactctgt gccagccaca caaacgttta cttatcttat gccacgggaa 4260
gtttagagag ctaagattat ctggggaaat caaaacaaaa aacaagcaaa caaaaaaaaa 4320
a 4321

<210> 21
<211> 919
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 21
Met Glu Val Gln Leu Gly Leu Gly Arg Val Tyr Pro Arg Pro Pro Ser
1 5 10 15
Lys Thr Tyr Arg Gly Ala Phe Gln Asn Leu Phe Gln Ser Val Arg Glu
20 25 30
Val Ile Gln Asn Pro Gly Pro Arg His Pro Glu Ala Ala Ser Ala Ala
35 40 45
Pro Pro Gly Ala Ser Leu Leu Leu Leu Gln Gln Gln Gln Gln Gln Gln
50 55 60
Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Glu Thr
65 70 75 80
Ser Pro Arg Gln Gln Gln Gln Gln Gly Glu Asp Gly Ser Pro Gln
85 90 95
Ala His Arg Arg Gly Pro Thr Gly Tyr Leu Val Leu Asp Glu Glu Gln
100 105 110
Gln Pro Ser Gln Pro Gln Ser Ala Leu Glu Cys His Pro Glu Arg Gly
115 120 125

```
Cys Val Pro Glu Pro Gly Ala Ala Val Ala Ala Ser Lys Gly Leu Pro
    130                 135                 140
Gln Gln Leu Pro Ala Pro Pro Asp Glu Asp Asp Ser Ala Ala Pro Ser
145                 150                 155                 160
Thr Leu Ser Leu Leu Gly Pro Thr Phe Pro Gly Leu Ser Ser Cys Ser
                165                 170                 175
Ala Asp Leu Lys Asp Ile Leu Ser Glu Ala Ser Thr Met Gln Leu Leu
            180                 185                 190
Gln Gln Gln Gln Gln Glu Ala Val Ser Glu Gly Ser Ser Ser Gly Arg
        195                 200                 205
Ala Arg Glu Ala Ser Gly Ala Pro Thr Ser Ser Lys Asp Asn Tyr Leu
    210                 215                 220
Gly Gly Thr Ser Thr Ile Ser Asp Asn Ala Lys Glu Leu Cys Lys Ala
225                 230                 235                 240
Val Ser Val Ser Met Gly Leu Gly Val Glu Ala Leu Glu His Leu Ser
                245                 250                 255
Pro Gly Glu Gln Leu Arg Gly Asp Cys Met Tyr Ala Pro Leu Leu Gly
            260                 265                 270
Val Pro Pro Ala Val Arg Pro Thr Pro Cys Ala Pro Leu Ala Glu Cys
        275                 280                 285
Lys Gly Ser Leu Leu Asp Asp Ser Ala Gly Lys Ser Thr Glu Asp Thr
    290                 295                 300
Ala Glu Tyr Ser Pro Phe Lys Gly Gly Tyr Thr Lys Gly Leu Glu Gly
305                 310                 315                 320
Glu Ser Leu Gly Cys Ser Gly Ser Ala Ala Ala Gly Ser Ser Gly Thr
                325                 330                 335
Leu Glu Leu Pro Ser Thr Leu Ser Leu Tyr Lys Ser Gly Ala Leu Asp
            340                 345                 350
Glu Ala Ala Ala Tyr Gln Ser Arg Asp Tyr Tyr Asn Phe Pro Leu Ala
        355                 360                 365
Leu Ala Gly Pro Pro Pro Pro Pro Pro Pro Pro His Pro His Ala Arg
    370                 375                 380
Ile Lys Leu Glu Asn Pro Leu Asp Tyr Gly Ser Ala Trp Ala Ala Ala
385                 390                 395                 400
Ala Ala Gln Cys Arg Tyr Gly Asp Leu Ala Ser Leu His Gly Ala Gly
                405                 410                 415
Ala Ala Gly Pro Gly Ser Gly Ser Pro Ser Ala Ala Ala Ser Ser Ser
            420                 425                 430
Trp His Thr Leu Phe Thr Ala Glu Glu Gly Gln Leu Tyr Gly Pro Cys
        435                 440                 445
Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
    450                 455                 460
Gly Gly Gly Gly Gly Gly Gly Gly Glu Ala Gly Ala Val Ala Pro Tyr
465                 470                 475                 480
Gly Tyr Thr Arg Pro Pro Gln Gly Leu Ala Gly Gln Glu Ser Asp Phe
                485                 490                 495
Thr Ala Pro Asp Val Trp Tyr Pro Gly Gly Met Val Ser Arg Val Pro
            500                 505                 510
Tyr Pro Ser Pro Thr Cys Val Lys Ser Glu Met Gly Pro Trp Met Asp
        515                 520                 525
Ser Tyr Ser Gly Pro Tyr Gly Asp Met Arg Leu Glu Thr Ala Arg Asp
    530                 535                 540
His Val Leu Pro Ile Asp Tyr Tyr Phe Pro Pro Gln Lys Thr Cys Leu
545                 550                 555                 560
Ile Cys Gly Asp Glu Ala Ser Gly Cys His Tyr Gly Ala Leu Thr Cys
                565                 570                 575
Gly Ser Cys Lys Val Phe Phe Lys Arg Ala Ala Glu Gly Lys Gln Lys
            580                 585                 590
Tyr Leu Cys Ala Ser Arg Asn Asp Cys Thr Ile Asp Lys Phe Arg Arg
        595                 600                 605
```

```
Lys Asn Cys Pro Ser Cys Arg Leu Arg Lys Cys Tyr Glu Ala Gly Met
    610                 615                 620
Thr Leu Gly Ala Arg Lys Leu Lys Lys Leu Gly Asn Leu Lys Leu Gln
625                 630                 635                 640
Glu Glu Gly Glu Ala Ser Ser Thr Thr Ser Pro Thr Glu Glu Thr Thr
                645                 650                 655
Gln Lys Leu Thr Val Ser His Ile Glu Gly Tyr Glu Cys Gln Pro Ile
            660                 665                 670
Phe Leu Asn Val Leu Glu Ala Ile Glu Pro Gly Val Val Cys Ala Gly
        675                 680                 685
His Asp Asn Asn Gln Pro Asp Ser Phe Ala Ala Leu Leu Ser Ser Leu
    690                 695                 700
Asn Glu Leu Gly Glu Arg Gln Leu Val His Val Val Lys Trp Ala Lys
705                 710                 715                 720
Ala Leu Pro Gly Phe Arg Asn Leu His Val Asp Asp Gln Met Ala Val
                725                 730                 735
Ile Gln Tyr Ser Trp Met Gly Leu Met Val Phe Ala Met Gly Trp Arg
            740                 745                 750
Ser Phe Thr Asn Val Asn Ser Arg Met Leu Tyr Phe Ala Pro Asp Leu
        755                 760                 765
Val Phe Asn Glu Tyr Arg Met His Lys Ser Arg Met Tyr Ser Gln Cys
    770                 775                 780
Val Arg Met Arg His Leu Ser Gln Glu Phe Gly Trp Leu Gln Ile Thr
785                 790                 795                 800
Pro Gln Glu Phe Leu Cys Met Lys Ala Leu Leu Leu Phe Ser Ile Ile
                805                 810                 815
Pro Val Asp Gly Leu Lys Asn Gln Lys Phe Phe Asp Glu Leu Arg Met
            820                 825                 830
Asn Tyr Ile Lys Glu Leu Asp Arg Ile Ile Ala Cys Lys Arg Lys Asn
        835                 840                 845
Pro Thr Ser Cys Ser Arg Arg Phe Tyr Gln Leu Thr Lys Leu Leu Asp
    850                 855                 860
Ser Val Gln Pro Ile Ala Arg Glu Leu His Gln Phe Thr Phe Asp Leu
865                 870                 875                 880
Leu Ile Lys Ser His Met Val Ser Val Asp Phe Pro Glu Met Met Ala
                885                 890                 895
Glu Ile Ile Ser Val Gln Val Pro Lys Ile Leu Ser Gly Lys Val Lys
            900                 905                 910
Pro Ile Tyr Phe His Thr Gln
        915
```

37 P 1 (1)

PATENT COOPERATION TREATY

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference	FOR FURTHER ACTION	see Notification of Transmittal of International Search Report (Form PCT/ISA/220) as well as, where applicable, item 5 below.
CG/12326.19		
International application No.	International filing date *(day/month/year)*	(Earliest) Priority Date *(day/month/year)*
PCT/CA 99/ 00852	15/09/1999	15/09/1998
Applicant		
SIGNALGENE INC. et al.		

This International Search Report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This International Search Report consists of a total of 3 sheets.

[X] It is also accompanied by a copy of each prior art document cited in this report.

1. **Basis of the report**

   a. With regard to the **language**, the international search was carried out on the basis of the international application in the language in which it was filed, unless otherwise indicated under this item.

   [ ] the international search was carried out on the basis of a translation of the international application furnished to this Authority (Rule 23.1(b)).

   b. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, the international search was carried out on the basis of the sequence listing :

   [ ] contained in the international application in written form.

   [ ] filed together with the international application in computer readable form.

   [X] furnished subsequently to this Authority in written form.

   [X] furnished subsequently to this Authority in computer readble form.

   [X] the statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.

   [X] the statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished

2. [ ] **Certain claims were found unsearchable** (See Box I).

3. [ ] **Unity of invention is lacking** (see Box II).

4. With regard to the **title**,

   [X] the text is approved as submitted by the applicant.

   [ ] the text has been established by this Authority to read as follows:

5. With regard to the **abstract**,

   [X] the text is approved as submitted by the applicant.

   [ ] the text has been established, according to Rule 38.2(b), by this Authority as it appears in Box III. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. The figure of the **drawings** to be published with the abstract is Figure No. —

   [ ] as suggested by the applicant.

   [ ] because the applicant failed to suggest a figure.

   [ ] because this figure better characterizes the invention.

   [ ] None of the figures.